Bounded Gaps Between Consecutive Primes

John Musson
Trinity College
MMath
Hilary 2015
Abstract

In 2013 Yitang Zhang proved a weak version of the Twin Prime Conjecture, stating that there are infinitely many pairs of consecutive prime numbers within a bounded distance from each other. Zhang’s bound was 70,000,000, but within a year James Maynard had produced a proof using significantly different methods that reduced this bound to 600. This dissertation will give an exposition of Maynard’s recent paper, in a way that will make it accessible to undergraduate or recently graduated number theorists.
## Contents

1 Introduction .................................................. 2

2 History ......................................................... 2

3 Things You Need to Know .................................... 9

3.1 Notation ..................................................... 9

3.2 Möbius Function ............................................ 10

3.3 Distribution of the primes ................................ 10

3.4 The Bombieri - Vinogradov Theorem ...................... 11

3.5 Tuples ....................................................... 13

3.6 Sieves ....................................................... 15

4 Choosing Sieve Weights ...................................... 16

4.1 A Heuristic .................................................. 16

4.2 Simple weightings .......................................... 17

4.2.1 Selberg weights .......................................... 18

5 Maynard’s Results ............................................ 19

5.1 The Big Four Theorems .................................... 19

5.2 Our three supports ......................................... 21

6 On the proof of Proposition 1 ............................ 25

6.1 An overview of the proof ................................ 25

6.2 \( S_1 \) ......................................................... 25

6.3 \( S_2 \) ......................................................... 38

7 The Lagrangian Method ..................................... 44

8 Smooth choice of \( y \) ......................................... 50

9 Finale .......................................................... 60

9.1 Final Proofs .................................................. 60

9.2 Acknowledgements .......................................... 64
1 Introduction

The prime numbers have fascinated mathematicians since antiquity: from Euclid’s proof of their boundlessness to the Riemann Hypothesis, they could consistently claim the title of the greatest enigma in mathematics. Like many of the best mathematical problems, the Twin Prime Conjecture is simple to state yet devilishly tricky to prove - in fact, it is still eluding proof.

Conjecture (Twin Prime Conjecture). There are infinitely many primes \( p \), such that \( p + 2 \) is also a prime.

Hundreds of thousands of twin primes have been found using computational methods; examples include the well known pairs \((3, 5)\) and \((11, 13)\) and the lesser known pair \((101999, 102001)\).

It is unclear when the notion of twin primes was first coined. Some attribute it to Euclid (which would make this one of mathematics’ oldest unsolved problems), however the earliest observance in mathematical literature, according to a history of the problem by Nazardonyavi [13], can be found in Polignac’s “Six propositions arithmologiques déduites du crible d’Ératosthène” [14], in which he proposes the more general conjecture:

Conjecture (Polignac, 1849). For any positive even number \( n \), there are infinitely many consecutive primes \( p \) and \( q \) such that \( q - p = n \).

With \( n = 2 \) this is exactly the Twin Prime Conjecture. Developments since 2013 have finally put us within touching distance of these conjectures: for the first time we have an unconditional, finite bound on the minimum difference between arbitrarily large consecutive primes. This dissertation will mostly be a detailed exposition of a number of recent theorems by Maynard [11]; however we’ll begin with a short history of work in this area before covering the mathematical techniques used in the proof and then continuing on to the theorems themselves.

2 History

1793 - Gauss

Considered by many to be the greatest mathematician of all time, Carl Friedrich Gauss made huge contributions to many areas of mathematics in his
lifetime. His *Disquisitiones Arithmeticae*, which he published aged only 24, changed number theory from a collection of results into a recognised branch of mathematics. Gauss solved the ancient problem of how to construct a 17-gon with just ruler and compass; introduced modular arithmetic and, after calculating a list of primes up to 3,000,000, conjectured what would become known as the prime number theorem. The following theorem was independently proposed by Legendre around the same time albeit with the more slowly converging estimate of $\frac{x}{\log x}$:

**Theorem (Prime Number Theorem).** Let $\pi(x)$ be the number of primes below $x$, and let

$$
\text{Li}(x) := \int_2^x \frac{1}{\log(t)} \, dt.
$$

Then

$$
\lim_{x \to \infty} \frac{\pi(x)}{\text{Li}(x)} = 1.
$$

Number theory really blossomed following Gauss. Throughout the 19th century mathematicians such as Galois and Dirichlet developed whole new subfields, making greater levels of rigour and abstraction commonplace. The crowning achievement of 19th century number theory however, was probably the introduction of complex analytical methods with the work of Riemann concerning his now infamous zeta function in 1859. The Generalised Riemann Hypothesis (GRH), stating that every zero of certain functions, like the Riemann zeta function, lies on the line $\text{Re}(x) = \frac{1}{2}$, has stood the test of time and remains one of the great open problems of modern mathematics. It was a result of work done by Riemann, followed by Hadamard and de la Vallée Poussin in 1896, that finally supplied a proof to Gauss’ then century old conjecture which created the Prime Number Theorem above.

**1965 - Bombieri - Vinogradov**

Jumping ahead to the middle of the next century, number theory had progressed significantly. While the GRH was (and still is) unreachable in its full capacity, many related results had come to light, one which is the Bombieri - Vinogradov Theorem. Although Bombieri and A. Vinogradov (there were two Vinogradovs in 20th century Russian mathematics, so it pays to be specific) never published a joint paper, each published slightly different results in the same year, a refinement of which is the Bombieri - Vinogradov
Theorem [21]. In general terms, the Bombieri - Vinogradov Theorem gives an upper bound for the average number of primes less than a given \( x \in \mathbb{N} \) in a particular residue class modulo \( q \leq Q \) for some \( Q \leq x \). Specifically, defining
\[
\theta(x; q, a) := \sum_{\substack{p \leq x \\text{prime} \\text{p mod } q \equiv a \mod q}} \log(p),
\]
a weighted sum for the number of primes less than \( x \) in the residue class \( a \) modulo \( q \), we have:

**Theorem (Bombieri - Vinogradov, 1965).** The following inequality holds for \( Q \leq x^{\frac{1}{2}} (\log x)^{-C} \) for some positive constant \( C \) with \( A \) a positive constant depending on \( C \):

\[
\sum_{q \leq Q} \max_{(a,q)=1} \left| \theta(x; q, a) - \frac{x}{\phi(q)} \right| \ll \frac{x}{(\log(x))^A}.
\]

We will give a more detailed explanation of the precise meaning and consequences of this theorem in Section 3.4.

The exponent of \( x \) for which this theorem holds is called the level of distribution of the primes. It is often referred to as \( \theta \), not to be confused with the function \( \theta(x; q, a) \).

There has been a lot of work attempting to extend \( \theta \) past \( \frac{1}{2} \), in particular to prove a conjecture of Elliott - Halberstam [2]:

**Conjecture (Elliott - Halberstam Conjecture, 1970).** The Bombieri - Vinogradov Theorem holds for all \( Q < x(\log x)^{-C} \).

However no proof is yet known.

One of the major ingredients of recent work extending the Bombieri - Vinogradov Theorem was improving the bound to \( x^{\frac{1}{2}+\omega} \) for an arbitrarily small \( \omega > 0 \), albeit with a modified version of the Bombieri - Vinogradov statement, which was in turn sufficient to give bounded gaps between prime numbers. On the other hand Friedlander and Granville [3] showed that the bound of \( x \) is maximal (in particular, the statement does not hold for \( Q \leq x(\log x)^{-C} \) for any fixed \( C \)). The minimum distance of bounded gaps is considerably reduced if one assumes the full Elliott - Halberstam conjecture, though the existence of such a minimum depends only on Bombieri - Vinogradov.
Goldston, Pintz and Yıldırım published a groundbreaking paper in 2005 [4], from which mathematicians believed bounded gaps would come very quickly, but that took another eight years. Henceforth, for brevity, we shall refer to the three authors (or their paper) simply as GPY. This paper gives two major results: the first proves a long-standing conjecture about the density of the primes, the second is another step on the long and, as yet, unfinished road towards the prime \( k \)-tuples conjecture.

Using Gauss’ estimate for the density of primes we anticipate “on average” the distance between two primes to be the logarithm of either, i.e.

\[
\frac{p_{n+1} - p_n}{\log p_n} = 1.
\]

But instead what we would like to study is the smallest gap between primes that occur in any interval, that is, the behaviour of \( \inf_{[N, 2N]} (p_{n+1} - p_n) \) as \( N \) tends to infinity.

**Theorem (GPY, 2005).** Let \( p_n \) be the \( n \)th prime number, then we have

\[
\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} = 0.
\]

This theorem demonstrates a slower than logarithmic growth for the minimum distance between consecutive primes, which made the case for it being bounded very optimistic indeed. Consequently, in the same paper, GPY proceeded to give a result ‘a hair’s-breadth’ from bounded gaps. This led to the following theorem:

**Theorem (GPY, 2005).** Suppose the primes have level of distribution \( \alpha > \frac{1}{2} \). Then there exists an explicit constant \( C(\alpha) \), depending only on \( \alpha \), such that any admissible \( k \)-tuple with \( k \geq C(\alpha) \) contains at least two primes infinitely often.

In particular if \( \alpha \geq 0.971 \) then

\[
\liminf_{n \to \infty} (p_{n+1} - p_n) \leq 16.
\]

This result doesn’t quite give bounded gaps because it relies on the level of distribution of the primes being greater than \( \frac{1}{2} \), which remains unproven.
The tools used by GPY were almost a decade old when their paper was published. Their input was to apply Heath-Brown’s generalization of the Selberg sieve (which we will define in the next section), for estimating the number of prime factors in any admissible tuple [8] (another definition that will be addressed in Section 3), to the task of estimating the existence of multiple primes in the tuple. The publication of this work prompted a great deal of mathematical activity attempting to remove the conditionality on an unproven statement. However, little progress was made for the next eight years.

2013 - Zhang

Yitang Zhang rocketed from obscurity as a lecturer at the University of New Hampshire to international fame (for a mathematician) at the remarkable age of 57 upon the publication of his paper providing the first proof of finite gaps between prime numbers [23].

Theorem (Zhang, 2013). Let $p_n$ be the $n^{th}$ prime number, it is proved that

$$\lim \inf_{n \to \infty} (p_{n+1} - p_n) \leq 70,000,000.$$  

This may seem a far cry from the twin primes conjecture, but 70,000,000 is considerably closer to 2 than infinity and his paper explicitly left significant room for others to sharpen his bounds.

The original part of Zhang’s proof was not to show that the primes have level of distribution $\alpha > \frac{1}{2}$ (which would have made the GPY method sufficient), but instead to use a stronger version of Bombieri - Vinogradov published by Bombieri - Friedlander - Iwaniec in the late 1980’s. Applicable only with an additional restriction on the prime factors of certain summands, many had thought it could be easily merged with GPY but each required conditions incompatible with the other. Following GPY, Iwaniec himself had tried, but failed, to remove the conditions on his paper. The most impressive thing about Zhang’s paper, was arguably how he managed to take the work of leading mathematicians in the field, and prove something with a method they had abandoned [22]. A great exposition of Zhang’s proof was written by Granville and can be found at [7].
As soon as Zhang’s work was available to the wider mathematical community it was devoured by some of the most prominent mathematicians in the area, and within two weeks the bound had been brought below 60,000,000 by Mark Lewko and then Thomas Trudgian [20]. The remarkable speed of progress and the potential for lowering such an important bound inspired mathematicians to work together to form a Polymath project.

The Polymath Project was conceived in 2009 by Tim Gowers [6] and Michael Nielsen as an experiment in massively collaborative online mathematics. The premise was that most mathematical collaborations are done between two, three or in rare cases four co-authors (with the notable exception of the Bourbaki group in the mid 1930s), but there are obvious advantages to having input from a larger pool of mathematicians; including generating a greater knowledge base, allowing people to focus on single parts of a work, and encouraging different approaches to the same problem. Polymath is still a work in progress and its projects have had varying levels of success. Of the nine full projects since its inception, six have produced partial answers or answers to related problems and three have created publishable results, including Polymath8, the follow up to Zhang’s work.

The Polymath8 project was headed up by Fields medalist Terence Tao on his blog [19]\(^1\) and the brief he gave was twofold: firstly to understand and clarify Zhang’s argument and secondly to improve the numerical bound on gaps between primes. The project was hugely successful, with the first write up totalling 163 pages and giving a reduction of Zhang’s bound to 4,680. But over the three months it was being written up (again, by a group of people, in the spirit of Polymath), a new paper came to light: written by James Maynard, the proofs involved quite different methods to Zhang and brought the upper bound down to 600 using the Bombieri - Vinogradov Theorem (not Zhang’s stronger alternative) and an improvement on GPY. While the write up of Polymath8 [15] was completed, Polymath8b was instantly spawned to see how their previous work could be applied to Maynard’s theorems to give an even stronger result. At the conclusion of Polymath8b, the best unconditional bound for the smallest gap between consecutive prime numbers is 246 [16].

\(^1\)Due to the nature of the blog style it is impossible to supply a single reference page, this reference gives all the pages related to Polymath8, however the reader is encouraged to browse T.Tao’s wordpress articles
Theorem (Polymath, 2014). Let $p_n$ be the $n$th prime number, unconditionally we have

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \leq 246.$$ 

They also gave a number of explicit bounds for $\liminf(p_{n+k} - p_n)$ for $1 \leq k \leq 5$ and a bound for general $k \in \mathbb{N}$ however we shall leave the more general cases in favour of the main focus of this paper.

2014 - Maynard

After finishing his DPhil at Oxford under Roger Heath-Brown, James Maynard took up a postdoctoral fellowship at the University of Montreal. While there, he finished a considerably more simple proof than Zhang’s of a finite bound for $\liminf(p_{n+1} - p_n)$, which also provided the significantly lower number of 600 mentioned above. Slightly weaker results were achieved independently by Terence Tao using similar methods at the same time causing some to refer to the “Maynard-Tao method”. This paper will go through Maynard’s proof and explain it at a level accessible to undergraduates or recent graduates in the area.

A Road Map for this Paper

In Section 3 we shall establish some notation as well as cover a few key mathematical tools for dealing with prime number theory. In Section 4 we shall look at some simple sieve weights and how we must adjust to get a sufficient level of complexity to produce interesting results. Section 5 will state the four theorems proven in Maynard’s paper, along with the propositions necessary for proving them. Sections 6,7 and 8 together will give us a proof of the main proposition. Finally Section 9 will pull the three propositions together to prove the four theorems as well as discuss further progress.
3 Things You Need to Know

The first thing we are going to do is establish the notation we are going to use throughout the paper and take some time explaining a few of them.

3.1 Notation

All variables $a, a_i, b, b_i, d, d_i, e, e_i, r, r_i, j, k, n, N, W$ and $x$ should be assumed to lie in the natural numbers unless explicitly specified otherwise.

$p$: a prime number

$p_n$: the $n^{th}$ prime number

$\epsilon, \delta$: any sufficiently small, positive number

$y_r$: $y_{r_1,...,r_k}$, a number depending on $r_1, \ldots, r_k$ and we will use these two notations interchangeably

$a \mid b$: $a$ divides $b$

$a \nmid b$: $a$ does not divide $b$

$a \equiv b (p)$: $a$ is congruent to $b$ modulo $p$

$\mu(x)$: The Möbius function (covered in Section 3.2)

$\varphi(x) = \prod_{p|x, p \text{ prime}} (p - 1)$

$g(x) = \prod_{p|x, p \text{ prime}} (p - 2)$

$\phi(x)$: the Euler Totient function $\phi(q) = |\{p < q : (p, q) = 1\}|$

$\theta(x)$: the first Chebyshev function, defined in Section 3.3

$\theta(x; q, a)$: defined in Section 3.3

$\tau_j(a)$: the divisor function (the number of ways of writing $a$ as a product of $j$ factors)

$\sum_{d_1, \ldots, d_k}$: the sum over all possible sets $\{d_1, \ldots, d_k\}$

$H = (h_1, \ldots, h_k)$: a $k$-tuple (covered in Section 3.4), assumed to be fixed

$(a, b)$: the highest common factor of $a$ and $b$ (at no point are $(a, b)$ and $(h_1, \ldots, h_k)$ used together so no confusion should arise)

$[a, b]$: the lowest common multiple of $a$ and $b$

$\lceil x \rceil$: the roof of $x$, the smallest integer greater than $x$

$O_A$: big $O$ asymptotic notation, depending on $A$. If no $A$ is specified then it may depend on $k$ and $H$
3.2 Möbius Function

The Möbius function is defined as follows:

\[
\mu(x) := \begin{cases} 
1 & x = 1 \\
1 & x \text{ is square-free and has an even number of prime factors} \\
-1 & x \text{ is square-free and has an odd number of prime factors} \\
0 & x \text{ is not square-free.}
\end{cases}
\]

It can be used as an indicator of whether a positive integer \( n \) is 1 or not:

\[
\sum_{d|n} \mu(d) = \begin{cases} 
1 & n = 1 \\
0 & n > 1
\end{cases}
\]

It is multiplicative for coprime numbers, that is, \( \mu(a)\mu(b) = \mu(ab) \). Clearly if \( a \) and \( b \) are not coprime then their product is not squarefree and in that case \( \mu(ab) = 0 \).

3.3 Distribution of the primes

Understanding fully the distribution of the primes is beyond current mathematics; however we have had estimates for how they propagate in general since the 18th century.

Let \( \pi(x) \) be the prime counting function that increments by 1 for every prime between 0 and \( x \). In 1798 Legendre [10] estimated that

\[
\frac{x}{\log(x) - 1.08366} \sim \pi(x).
\]

Almost 50 years later, Dirichlet (in communication with Gauss) used his integral

\[
\text{Li}(x) := \int_2^x \frac{1}{\log(t)} \, dt
\]

for the same approximation. The difference between \( \text{Li}(x) \) and \( \pi(x) \) is much smaller than Legendre’s function; however convergence in both cases is equivalent when it comes to the quotient. This means the Prime Number Theorem can be stated as either:

\[
\lim_{x \to \infty} \frac{\pi(x)}{\frac{x}{\log(x)}} = 1,
\]
or

\[
\lim_{x \to \infty} \frac{\pi(x)}{\text{Li}(x)} = 1.
\]

We define the Chebyshev function, a weighted prime counting function,

\[
\theta(x) := \sum_{\substack{p \leq x \\ p \text{ prime}}} \log(p)
\]

and, as mentioned previously,

\[
\theta(x; q, a) := \sum_{\substack{\text{p} \leq x \\ \text{p prime} \\ p \equiv a \ (q)}} \log(p).
\]

It turns out that the Prime Number Theorem is equivalent to the statement

\[
\frac{\theta(x)}{x} \sim 1 \quad \text{as } x \to \infty
\]

(a proof of this can be found in [9] page 83, albeit in German).

### 3.4 The Bombieri - Vinogradov Theorem

We now provide the motivation behind the Bombieri - Vinogradov Theorem which we stated in the previous section. We know that being able to estimate \( \pi(x) \) is equivalent to being able to estimate \( \theta(x) \), but unfortunately both are rather hard to do. Instead we have the Bombieri - Vinogradov Theorem which we now have all the tools necessary to understand. The precise nature of the function used in the Bombieri - Vinogradov Theorem takes a little explaining:

1. First, consider how we anticipate the primes below \( x \) to be distributed modulo \( q \) for any given \( q \). If the primes were a totally random distribution we could conjecture that \( \{|p \leq x : p = a \mod q\| = \frac{\pi(x)}{\phi(q)} \) where \( \phi(q) \) is the Euler totient function \( (\phi(q) = |\{p < q : (p, q) = 1\}|) \);

2. From the Prime Number Theorem we have \( \pi(x) \to \frac{x}{\log(x)} \);

3. Finally, the weighting of each prime from 1 to \( \log(p) \) in the definition of \( \theta \) gives us \( \theta(x; q, a) = \frac{\pi(x)}{\phi(q)} \log(x) \approx \frac{x}{\phi(x)} \).
We do in fact have results saying exactly this, but the nature of the relation $\approx$ naturally leaves us with errors to minimise. Let’s have a look at how small we can make these errors:

The prime number theorem can be extended to a version for arithmetic progressions in the same way that $\theta(x)$ is extended to $\theta(x; q, a)$. From this we get the following (see [12, Corollary 11.21] for a proof):

**Corollary 3.1.** Let $x, q, a \in \mathbb{N}$ and $A > 0$ be such that $x$ and $q$ are coprime and $q \leq \log^A x$, then

$$\theta(x; q, a) = \frac{x}{\phi(q)} + O_A(x(\log x)^{-A}).$$

If, however, we allow ourselves to assume the full Generalised Riemann Hypothesis, we get a much stronger bound [12, Corollary 13.8].

**Corollary 3.2.** Assume the Generalised Riemann Hypothesis and let $x, q, a, \in \mathbb{N}$ and $A > 0$ be such that $x$ and $q$ are coprime and $q \leq \log^A x$, then

$$\theta(x; q, a) = \frac{x}{\phi(q)} + O_A\left(x^{\frac{1}{2}}(\log x)^{2}\right).$$

Unfortunately we can’t just assume the GRH is true, but luckily for us, Bombieri - Vinogradov is a completely unconditional, yet very strong, result. What this gives us is that ‘on average’ Corollary 3.2 is true (without first assuming GRH); in this case, by ‘on average’ we mean that if we take the sum over all $q \leq x^m$, we expect the error to be no larger than $O_A \left(x^m x^{\frac{1}{2}}(\log x)^{2}\right)$.

We must note, however, that Bombieri - Vinogradov extends only to $m < \frac{1}{2}$.

For each $q$, Bombieri - Vinogradov takes the worst case, the residue class $a$ with either the most or the fewest primes (depending on which is further from $\frac{x}{\phi(q)}$) and we sum over all $q \leq \mathcal{Q}$; this allows us a little leeway, for some $q$ it may be the case that

$$\max_{(a,q)=1} |\theta(x; q, a) - \frac{x}{\phi(q)}| > \frac{x}{(\log(x)^A \mathcal{Q})} > \frac{x^{\frac{1}{2}}}{\log(x)^A}. $$

That is, the difference on the LHS is greater than the magnitude of error given in Corollary 3.2, but the theorem tells us that this is not the case in general.

The statement is as follows [21]:
Theorem (Bombieri - Vinogradov, 1965). For all $x$ in $\mathbb{N}$, and all positive constants $A$ and $C = f(A)$, the following holds if $Q < x^{1/2}(\log x)^{-C}$:

$$\sum_{q \leq Q} \max_{(a,q)=1} |\theta(x; q, a) - \frac{x}{\phi(q)}| \ll \frac{x}{(\log x)^A}.$$ 

Finally, for completeness, we restate that if Bombieri - Vinogradov holds for $Q < x^{\theta}(\log x)^{-C}$ for some constant $C \in \mathbb{R}$, then $\theta$ is the level of distribution of the primes, and we give the extension of this result (conjectured a few years later) that has been left to future mathematicians to prove [2]:

Conjecture (Elliott - Halberstam Conjecture, 1970). The Bombieri - Vinogradov Theorem holds for all $Q < x(\log x)^{-C}$.

We include this again because, once proved, it will overtake the Bombieri - Vinogradov as the prominent result in the area. For this reason, most papers on bounding prime numbers, and certainly all those referenced in this dissertation, give two results, one assuming Elliott - Halberstam and one attainable with only current methods.

3.5 Tuples

The Bombieri - Vinogradov Theorem concerns arithmetic progressions so we now develop some terminology to represent different progressions.

Consider an ordered set of distinct, non-negative integers, for example $H = (0, 1, 5)$. Then the corresponding tuples are all sets \{n, n + 1, n + 5\}, (n $\in$ $\mathbb{N}$), so our $H$ stands for all the sets of integers with the same additive pattern. A $k$-tuple helpfully contains $k$ elements and we have a prime $k$-tuple when every element is prime. The length of a tuple is the difference between the first and last element.

A more relevant example would be the twin primes, which are represented by the tuple (0, 2): the Twin Prime Conjecture is exactly the statement that this tuple corresponds to infinitely many prime 2-tuples.

Conjecture (Prime $k$-tuples Conjecture). For any admissible prime $k$-tuple, $(h_1, h_2, \ldots, h_k)$, the numbers \{n + h_i\}_{i=1}^k are all prime for infinitely many $n$ in $\mathbb{N}$.
A tuple \( H = (h_1, h_2, \ldots, h_k) \) is ‘admissible’ if, for every prime \( p \), there is some integer \( a_p \) less than \( p \) such that no \( h_n \) in \( H \) is congruent to \( a_p \) modulo \( p \). That is, if you reduced every element of \( H \) to its residue mod \( p \), at least one number from 0 to \( p - 1 \) is not in \( H \); equivalently, any set corresponding to \( H \) may be referred to as an ‘admissible’ set. For every prime \( p \) greater than \( k \), this must be the case as there are more residue classes than elements in \( H \). So for any finite \( H \), whether it is admissible or not is a finite check. A set is inadmissible if it is not admissible.

Being admissible is important because only admissible tuples have a chance of representing infinitely many prime \( k \)-tuples. Looking back to our original \( H \) we see it contains an element from every residue class modulo 3. Considering the general set \( \{n, n + 1, n + 5\} \) we note if \( n \) is prime it is either 2, or odd, in which case \( n + 5 \) is even (and not 2) so not prime, hence the only prime tuple represented by \( H \) is \{2, 3, 7\} and \( H \) becomes somewhat redundant in the search for infinite sets of primes with specific differences. This motivates our first proposition:

**Proposition 3.1.** For an inadmissible tuple there are at most \( k \) corresponding prime \( k \)-tuples.

*Proof.* Let \( H = (h_1, h_2, \ldots, h_k) \) be an inadmissible tuple, ordered so that \( h_{n-1} < h_n \) for every \( n \leq k \). Without loss of generality we may let \( h_1 \) be the smallest number such that all the \( h_i \) are prime, recall we are only interested in the differences \( h_i - h_1 \) in our definition of \( H \). If no such \( h_1 \) exists then there are no corresponding prime \( k \)-tuples and we are done.

By definition of inadmissible, for some prime \( p < k \), \( H \) contains every residue class of \( p \). Specifically, \( H \) always contains an element which is congruent to 0 mod \( p \). That is, \( H \) contains an element which is a multiple of \( p \). Either it is \( p \), or it is not prime.

As \( \{h_1, h_2, \ldots, h_k\} \) contains only primes, then from above it contains \( p \), say \( h_t \). As elements in \( H \) are in increasing order, if, for some integer \( n \), \( n + h_s \equiv 0 (p) \) for any \( s > t \) then \( n + h_s \) must be greater than \( p \) and thus be a multiple of \( p \) and clearly not be prime. Hence, there are at most \( t - 1 \) integers greater than 0 that may produce a set containing only primes, namely \((p - h_{t-1}), (p - h_{t-2}), \ldots, (p - h_1)\). As \( t \) is arbitrary but bounded above by \( k \), we get the lemma. \( \Box \)

The idea of ‘twin primes’ or even consecutive primes does not fundamentally rely on tuples, because we’re only looking for two primes, not
Working with tuples gives us an instant generalisation of our theorems about consecutive primes to theorems about, for example, $p_n - p_{n+k}$.

### 3.6 Sieves

When a mathematician begins to learn about sieve theory they are usually given the very unfortunate starting example of the sieve of Eratosthenes and a metaphor about pasta. It is unfortunate in two ways, firstly because it doesn’t do what modern sieves do and secondly because it gives unrealistic expectations about the power of the tool.

The sieve of Eratosthenes is a simple method for finding primes below a natural number $n$:

1. Write down the first $n$ natural numbers,
2. Cross off the number 1,
3. Write the first uncrossed number on a separate list,
4. Cross off every multiple of that number,
5. Repeat steps 3 and 4 until every number is crossed off.

The first time you do step 3 you put the number 2 on your list, then cross off all the multiples of 2, next you will write down 3 and remove all of its products and so on. The metaphor will tell the student who has just successfully found all the primes up to $n$ that they have taken their initial list, a pan full of pasta (primes) that they want, and the water (composite numbers) that it’s cooked in and poured it through a sieve leaving them with only a tasty meal; this is a rather unrealistic view of what modern sieve theory can achieve.

The sieve of Eratosthenes is very good at its job; the problem is that to apply it to an arbitrarily large set of numbers takes an equally large amount of time and, as our $n$ tends to infinity, Eratosthenes’ sieve falls by the wayside somewhat. Our focus will be on the Selberg sieve (also known as the (Selberg) upper bound sieve, or the (Selberg) $\Lambda^2$-sieve), a tool created in 1947 by Atle Selberg. It differs from the sieve of Eratosthenes in that, instead of telling you exactly which numbers are prime, given a set $[N, 2N]$ it gives a good estimate for how many are prime. Continuing our pasta metaphor, the Selberg sieve doesn’t remove all the water from our saucepan but instead
gives you a good estimate for how much is left. We will give a more detailed description of how the sieve works in the next section. Selberg used his sieve to give an upper bound for the number of primes remaining, but at the time it didn’t produce the level of results he had been looking for and was later replaced in most mathematicians’ minds by more sophisticated methods. So much so, in fact, that in 1991 [18] Selberg described it as “now of historic interest only”. However he was shortly proven wrong - just before the turn of the millennium, Roger Heath-Brown used a generalisation of the Selberg sieve in the search for prime tuples and this was later built on by GPY to set the scene for Zhang’s work.

4 Choosing Sieve Weights

4.1 A Heuristic

Assume we have an admissible set $H = (h_1, h_2, \ldots, h_k)$. Then we want to find $n$ such that ‘many’ of the \{n + h_i\}_{i=1}^k are prime.

We want to see how the number of primes in \{n + h_i\}_{i=1}^k changes as $n \to \infty$, so define

$$S := \sum_{n < N} \left| \left\{ p: p \in \{ n + h_i \}_{i=1}^k \right\} \right| w_n$$

for some, as yet undefined, weights $w_n \geq 0$.

We begin with a look at possible sieve weights and discuss limitations of the most obvious choices to motivate the weights we will eventually choose.

A useful tool for this is the prime indicator function

$$1_p(n) := \begin{cases} 1, & n \text{ prime} \\ 0, & \text{otherwise} \end{cases}$$

and we consider the sum

$$P_n := \sum_{i=1}^k 1_p(n + h_i)$$

which counts the number of primes for each $n$. It may seem odd that we go from considering $P_n$, which is precisely the value we want to understand, to $S$, a seemingly more complicated expression. This is because $P_n$ is “easy” to
understand only in a tautological sense: if we understood it or could evaluate it well, then we would know the truth of the prime k-tuples conjecture or, at the very least, bounded gaps. Instead we work with the sum $S$ which could be considered a weighted average over all the $P_n$ in a given range.

We can then redefine $S$ from the plan to be:

$$S := \sum_{N \leq n \leq 2N} (P_n - m)w_n. \quad (4.1)$$

The keen-eyed reader may notice that $n$ in our sum no longer runs from 0 up to $N$, but instead from $N$ to $2N$. We make this change because if we went from 0, we would have to keep changing our sieve weights to deal with all the smaller primes, so this simplifies things by giving our $n$ a lower bound. We will in fact use an $S$ different again from either of these in the forthcoming proofs, summing instead over $n \nmid W$ (where $W = \prod_{p < D_0} p$ is the product over all primes less than a certain constant), but these have the same effect and, to show the different options, we will continue to use (4.1) for now.

We want these weights to be large when $P_n > m$ and small otherwise. If $S > 0$ as $x \to \infty$ then at least one of the $P_n$ must be making a positive contribution, so there must be at least $m$ primes in the bounded interval $[n + h_1, n + h_k]$ for infinitely many $n$, $m = 2$ would be sufficient to give us bounded gaps, however it is good to work in slightly more generality. This leads us on to phase 2:

### 4.2 Simple weightings

Currently the only restriction we have on our weights is that they must be non-negative, but clearly this leaves rather a large range of options for our $w_n$; some of these will be better than others. Let’s begin with two trivial options: the simplest, and the most efficient.

1) The simplest case is when we add no weighting, or equivalently, a uniform weighting. That is, we take $w_n = \frac{1}{N}$ for all $n$. Then

$$S = \sum_{N \leq n \leq 2N} (P_n - m)w_n$$

$$= \frac{1}{N} \sum_{N \leq n \leq 2N} (P_n - m)$$

17
by the prime number theorem, \( P_n \to \frac{k}{\log(N)} \) as \( N \to \infty \), giving,
\[
\frac{1}{N} \left( \frac{Nk}{\log(N)} - Nm \right) = \frac{k}{\log(N)} - m
\]

But,
\[
\lim_{x \to \infty} \left( \frac{k}{\log(N)} - m \right) \leq 0
\]
so (thankfully, else this dissertation would be cut rather short) we cannot get the relation we want \((S > 0)\) without weights.

2) The other obvious and most optimistic case is to take
\[
w_n = \begin{cases} 
1 & P_n > m \\
0 & \text{otherwise}
\end{cases}
\]
If we could evaluate \( S \) with this weighting then we would immediately have bounded gaps. However, as before when we had to move away from evaluating \( P_n \) on its own, claiming we knew enough about \( w_n \) to work with \( S \) would also be equivalent to knowing the proof of the statement in advance.

4.2.1 Selberg weights

Having looked at the two most extreme cases, we look now for one that sits in the middle. The Selberg sieve weights offer something we can evaluate and will eventually come to modify to get the results we want.

Selberg’s initial work was focused on looking at just 2-tuples and, on the face of it, this seems like the best way to find bounded gaps, seeing as a limit on finite gaps only requires two primes to be found. However, work since (in particular, the work we shall cover here) has looked at \( k \)-tuples because they offer us a lot more flexibility on where the two primes may lie.

With this in mind we’ll jump immediately to the previously mentioned \( k \)-dimensional generalisation of the Selberg sieve developed by Heath-Brown in 1997 [8]. Heath-Brown wanted to study a function very similar to our \( S \) above and so tried to find a weighting where the negative part is small and
the positive part is calculable. This led him to use the upper bound sieve to evaluate the negative part, with the following weighting:

\[ w_n := \left( \sum_{d \mid \prod_{i=1}^{k} (n+h_i)} \lambda_d \right)^2 \]

where

\[ \lambda_d \approx \begin{cases} 
\mu(d) \log^k \left( \frac{x}{d} \right), & \text{for all } d \leq \xi \\
0, & \text{for all } d > \xi 
\end{cases} \]

with \( \xi \) here being \( x^{\frac{3}{2} - \epsilon} \) for a small positive \( \epsilon \). This was sufficient for the results Heath-Brown was looking for but not enough to give bounded gaps. He notes in his paper that part of his analysis would break down if he took the exponent of \( x \) to be \( \frac{1}{2} - \epsilon \) and, as a result, his argument is somewhat inefficient. On the other hand, he also says that Selberg suggests replacing

\[ \sum_{d \mid \prod_{i=1}^{k} (n+h_i)} \lambda_d \]

with

\[ \sum_{d_1 \mid n+h_i} \lambda_{d_1, \ldots, d_k} \]

and this is exactly what Maynard did to gain the flexibility to establish much stronger results.

5 Maynard’s Results

James Maynard produced four theorems, covering both the general case of prime \( k \)-tuples and the specific case of consecutive primes, both with and without Elliott - Halberstam. The theorems fall into place relatively easily after we establish the corresponding propositions which will take up the majority of the remainder of this paper.

5.1 The Big Four Theorems

Theorem 1 is the specific and unconditional case that we are most interested in: the case of consecutive primes given only \( \theta < \frac{1}{2} \) (recall: \( \theta \) is the “level of
distribution of the primes” defined in Section 3.4). In Maynard’s paper it is given as 600, though he mentions this is not optimal, and since then there have been refinements by Tao, Kowalski and others in the Polymath project to bring this number down to 246 [16].

**Theorem 1.** Let $p_n$ be the $n^{th}$ prime, then

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \leq 600.$$ 

Similar to many other papers in this area, we provide a further bound of the type of Theorem 1 with the additional assumption of the Elliott - Halberstam conjecture.

**Theorem 2.** Assuming Elliott - Halberstam, we have:

$$\liminf_{n \to \infty} p_{n+1} - p_n \leq 12.$$ 

Theorem 3 gives a limit for a more general version of Theorem 1. Instead of looking for consecutive primes, it bounds the minimum gap of $p_{n+m} - p_n$. That is, for every $m$, there exists a constant $C_m$, such that there are infinitely many gaps of length $C_m$ containing at least $m$ primes. The bound predicted by the prime number theorem is $C_m \approx m \log m$, but the bound given in the paper is much weaker, specifically:

**Theorem 3.** Let $p_n$ be the $n^{th}$ prime number, $m \in \mathbb{N}$, then we have

$$\liminf_{n \to \infty} (p_{n+m} - p_n) \ll m^3 e^{4m}.$$ 

Finally, Theorem 4 offers a partial result related to the Prime $k$-tuples Conjecture; it says that for any admissible set $\mathcal{A}$, a non-trivial number of its subsets satisfy the condition in the Prime $k$-tuples Conjecture. This does not amount to the whole conjecture, as it does not prove it for all admissible tuples. Though we do anticipate equality in place of the $\gg$ symbol. Maynard’s original theorem allows $\mathcal{A}$ to be any set, but we state a slightly weaker version which has a somewhat simpler proof.

**Theorem 4.** Let $\mathcal{A} = \{a_1, a_2, \ldots, a_r\}$ be an admissible set of distinct integers and $H = \{h_1, h_2, \ldots, h_k\}$ be a subset of $\mathcal{A}$ containing $k < r$ elements. Then for a positive proportion of such $H$, $\{n + h_i\}_{i=1}^k$ are all prime for infinitely many $n$. 

20
Equivalently,

\[
\frac{|\{H \subseteq A : H \text{ satisfies the prime } k\text{-tuples conjecture}\}|}{|\{H \subseteq A\}|} \gg 1.
\]

We make a small note that the notation \(\gg\) does not imply the fraction is greater than 1, as this is impossible, but instead that it is greater than some unspecified positive constant.

These four theorems will follow easily from the remaining three propositions. The majority of the remainder of this paper will be taken up with proving Proposition 1. Proposition 2 shall be proved shortly after its statement, and for Proposition 3 we will refer the reader to Maynard’s paper due to length restrictions.

5.2 Our three supports

Recall from (4.1) the sum

\[S = \sum_{N \leq n \leq 2N} (P_n - m)w_n,\]

which is positive (with the correct weights \(w_n\)) if there is an \(n\) in the interval \([N, 2N]\) such that \(\{n + h_i\}_{i=1}^k\) contains \(m\) primes. Recall also that we defined

\[w_n := \left( \sum_{d_i | n+h_i} \lambda_{d_1,\ldots,d_k} \right)^2.
\]

We will want to have the definitions

\[D_0 := \log \log \log N,\]

a number small relative to \(N\), but nevertheless increasing with \(N\), and

\[W := \prod_{p < D_0} p.
\]

This will have multiple uses, the most common being that when we set a number \(n\) to be coprime to \(W\), we have given \(n\) some useful properties. Firstly, and most obviously, \(n > W\), perhaps most helpfully though, \(n\) has
no “small” prime factors which means that we can estimate $\varphi(n) \approx n$. This is because of Euler’s product formula, which states

$$\varphi(n) = n \prod_{p \mid n} \left(1 - \frac{1}{p}\right),$$

and the smallest possible factor in this product is then of size $\left(1 - \frac{1}{\log \log \log x}\right)$. As our $x$ is going to be pretty large, this stays fairly close to 1.

We now add an addendum to our definition of $w_n$: that $w_n$ shall be 0 unless $n \equiv v \pmod{W}$. As we’re only looking for $S$ to be positive to show we have an appropriate $n$; if it is positive on this subset of $[N, 2N]$, then that is sufficient.

Next we split $S$ with our extra condition into two sums:

$$s_1 = \sum_{N \leq n \leq 2N}^{n \equiv v \pmod{W}} w_n \quad (5.1)$$

$$s_2 = \sum_{N \leq n \leq 2N}^{n \equiv v \pmod{W}} P_n w_n. \quad (5.2)$$

So $S = s_2 - ms_1$ is positive if $s_2$ is large in comparison to $s_1$, a fact we will prove in Section 8.

With this in place we can state the three propositions upon which our theorems rest.

For the first proposition we will define some notation to neaten the results. Let

$$I_k(F) = \int_0^1 \cdots \int_0^1 F(t_1, \ldots, t_k)^2 \, dt_1 \cdots dt_k$$

and let $J_k^{(m)}(F)$ be defined as the same integral with the function $F$ replaced by $\int_0^1 F(t_1, \ldots, t_k) dt_m$ and then subsequently not integrating with respect to $t_m$, that is,

$$J_k^{(m)}(F) = \int_0^1 \cdots \int_0^1 \left( \int_0^1 F(t_1, \ldots, t_k) \, dt_m \right)^2 \, dt_1 \cdots dt_{m-1} dt_{m+1} \cdots dt_k.$$
Proposition 1. Let $\delta > 0$ be small and $R = N^{\frac{\theta}{2} - \delta}$. Let $F$ be a piecewise differentiable function supported on \[ \left\{ (x_1, x_2, \ldots, x_k) : \sum_{i=1}^{k} x_i \leq 1 \right\} \] and $d_i \in \mathbb{N}$. We define $w_n$ by

\[
 w_n := \lambda_{d_1, \ldots, d_k} = \left( \prod_{i=1}^{k} \mu(d_i) \right) \sum_{r_1, \ldots, r_k \atop (r_i | r_i, W) = 1} \frac{\mu(\prod_{i=1}^{k} r_i)}{\prod_{i=1}^{k} \varphi(r_i)} F \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)
\]

if $(\prod_{i=1}^{k} d_i, W) = 1$ (where, $(a, b)$ denotes the highest common factor of $a$ and $b$) and 0 otherwise. Then we get, whenever $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$,

\[
 S_1 = (1 + o(1)) \frac{\varphi(W)^k N(\log R)^k}{W^{k+1}} I_k(F),
\]

\[
 S_2 = (1 + o(1)) \frac{\varphi(W)^k N(\log R)^k+1}{W^{k+1} \log N} \sum_{m=1}^{k} J_k^{(m)}(F).
\]

Proposition 2 asserts that, for any appropriate $F$, at least

\[
 r_k = \left\lceil \frac{\theta \sum_{m=1}^{k} J_k^{(m)}(F)}{2 I_k(F)} \right\rceil
\]

of the numbers in $H$ are prime for infinitely many $n$. That is, if taking the $F$ which maximises it gives $r_k \geq 2$, we deduce bounded gaps.

Proposition 2. Let the primes have level of distribution $\theta > 0$. Let $\delta > 0$ and $H$ be an admissible set. Let $I_k(F)$ and $J_k^{(m)}(F)$ be as in Proposition 1 and $S_k$ be the set of all functions that satisfy the definition given for $F$ in Proposition 1. Taking

\[
 M_k = \sup_{F \in S_k} \frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)}
\]

or, equivalently,

\[
 M_k = \sup_{F \in S_k} \frac{S_1 \log N}{S_2 \log R},
\]

23
then $r_k$ above becomes $\left\lceil \frac{\theta}{2} M_k \right\rceil$ and we have at least $r_k$ of $n + h_i$ are prime. Explicitly

$$
\lim_{n \to \infty} \inf (p_{n + r_k} - p_n) \leq \max_{1 \leq i \neq j \leq k} (h_i - h_j).
$$

Our final proposition gives two explicit results for $M_k$ and a general bound.

**Proposition 3.** Let $M_k$ be as in Proposition 2. We have:

1. $2 < M_5$
2. $4 < M_{105}$
3. $\log k - 2 \log \log k - 2 < M_k$ for sufficiently large $k$

We prove Proposition 2, assuming Proposition 1:

**Proof.** By the definition of $M_k$, there exists some $F_0 \in S_k$ such that

$$
\sum_{1}^{k} J_k^{(m)}(F_0) > (M_k - \epsilon)I_k(F_0).
$$

As we said at the start of this section we’re looking to make $S = S_2 - mS_1 > 0$. By Proposition 1:

$$
S = (1 + o(1)) \frac{\varphi(W)^k N \log R}{W^{k+1}} \left( \frac{\log R}{\log N} \sum_{m=1}^{k} J_k^{(m)}(F_0) - mI_k(F_0) \right)
$$

$$
= \frac{\varphi(W)^k N \log R}{W^{k+1}} \left( \frac{\log R}{\log N} \sum_{m=1}^{k} J_k^{(m)}(F_0) - mI_k(F_0) + o(1) \right),
$$

substituting $M_k$ into $S$ gives

$$
\geq \frac{\varphi(W)^k N \log R}{W^{k+1}} I_k(F_0) \left( \left( \frac{\theta}{2} - \epsilon \right) (M_k - \epsilon) - m + o(1) \right)
$$

$$
\geq C_N \left( \frac{\theta M_k}{2} - \epsilon \left( \frac{\theta}{2} + M_k - \epsilon \right) - \frac{\theta M_k}{2} + \delta + o(1) \right)
$$

$$
= C_N \left( \epsilon^2 - \epsilon \frac{\theta}{2} - \epsilon M_k + \delta + o(1) \right).
$$

24
Then choosing \( \epsilon < \frac{1}{2} \max(\frac{\delta}{M_k}) \) gives \( S > 0 \) for all large \( N \). As mentioned earlier, this means we have \( \lfloor m + 1 \rfloor \) or more primes in \( \{ h_i + n \}_{i=1}^k \).

If also \( \delta \) is less than the fractional part of \( \frac{\theta M_k}{2^x} \), then

\[
\lfloor m + 1 \rfloor = \left\lfloor \frac{\theta M_k}{2} \right\rfloor = r_k
\]

\[ \square \]

The proof of Proposition 3 is an optimisation problem that will be omitted in this paper due to length restrictions, however, it can be found in [11, Sections 7 & 8]. The next three sections of our work will prove Proposition 1, from which we may then deduce the four theorems above.

6 On the proof of Proposition 1

Proposition 1 is really the crux upon which all the theorems lie: it says, if we choose these given weights, then we can evaluate \( S_1 \) and \( S_2 \) to be certain sums which Propositions 2 and 3 tell us are sufficient for bounded gaps. We will be following the proof set out by Maynard, adding as much extra working and explanation as possible to hopefully improve clarity for the reader.

6.1 An overview of the proof

1. Split the sum into two manageable parts (5.1 and 5.2)
2. Choose \( w_n \) sieve weights based on the Selberg sieve (Section 4.2.1)
3. Change variable in each part to get a smoothness condition (Lemmas 6.1 and 6.3)
4. Optimise with Lagrangian multipliers (Section 7)

6.2 \( S_1 \)

We defined our \( \lambda_{d_1, \ldots, d_k} \) to fit with the Selberg \( \Lambda^2 \)-sieve; however this is not optimal for evaluating our sums, so our first lemma gives us a change of variables that makes this much easier. We take \( \lambda_{d_1, \ldots, d_k} \) to be non-zero only
on tuples \((d_1, \ldots, d_k)\) for which \(d = \prod d_i < R (= N^{\frac{\theta}{2} - \delta})\), \((d, W) = 1\) and \(\mu(d)^2 = 1\) (i.e. \(d\) squarefree, which in turn gives \((d_i, d_j) = 1\) whenever \(i \neq j\)).

**Lemma 6.1.** Let

\[
y_{r_1, \ldots, r_k} = \left(\prod_{i=1}^{k} \mu(r_i) \varphi(r_i)\right) \sum_{d_1, \ldots, d_k \mid r_i, d_i} \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} d_i}
\]

(6.1)

and

\[
y_{\text{max}} = \sup_{r_1, \ldots, r_k} |y_{r_1, \ldots, r_k}|.
\]

Then

\[
S_1 = \frac{N}{W} \sum_{r_1, \ldots, r_k} \frac{y_{r_1, \ldots, r_k}^2}{\prod_{i=1}^{k} \varphi(r_i)} + O \left(\frac{y_{\text{max}}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0}\right).
\]

**Proof.** We have

\[
S_1 = \sum_{\substack{N \leq n \leq 2N \\ n \equiv v(W)}} \left(\sum_{d_i \mid n + h_i} \lambda_{d_1, \ldots, d_k}\right)^2
\]

\[
= \sum_{d_1, \ldots, d_k} \lambda_{d_1, \ldots, d_k} \sum_{e_1, \ldots, e_k} \lambda_{e_1, \ldots, e_k} \sum_{\substack{N \leq n \leq 2N \\ n \equiv v(W) \mod [d_i, e_i] \mid n + h_i}} 1.
\]

(6.3)

See that, without the second sum in (6.3), the first runs over all possible divisors of all \(n + h_i\) (not even just \(n \in [N, 2N]\)). In particular, if \(d_i \mid n + h_i\) and \(e_i \mid m + h_i\) for all \(i\), we have no way of guaranteeing \(m = n\), so we use the condition \([d_i, e_i] \mid n + h_i\) in the second sum to ensure that the \(\lambda_{d_1, \ldots, d_k}\) and \(\lambda_{e_1, \ldots, e_k}\) are both components of the same \(w_n\). It serves to move the \(d_i \mid n + h_i\) to the sum not containing the \(\lambda\)s.

It is obvious that

\[
\sum_{\substack{N \leq n \leq 2N \\ n \equiv v(W)}} 1 = \frac{N}{W} + O(1).
\]

Then, as long as the \([d_i, e_i]\) are pairwise coprime, (a result of the \(\mu(d)^2 = 1\) condition we imposed on \(\lambda\)) we have by the Chinese Remainder Theorem,
for some $v' \in \{1, \ldots, W \prod_{i=1}^{k} [d_i, e_i]\}$:

$$
\sum_{N \leq n \leq 2N} 1 = \sum_{\substack{n \equiv v(W) \\ [d_i, e_i]}} 1 = \frac{N}{W \prod_{i=1}^{k} [d_i, e_i]} + O(1), \quad (6.4)
$$

giving us

$$
S_1 = \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k \atop W, [d_1, e_1], \ldots, [d_k, e_k]} \prod_{i=1}^{k} \lambda_{d_i} \lambda_{e_i} \left( \frac{N}{W \prod_{i=1}^{k} [d_i, e_i]} + O(1) \right). \quad (6.5)
$$

Simply for ease of notation, we let $\tilde{\sum}$ denote summation with the restriction $W, [d_1, e_1], \ldots, [d_k, e_k]$ pairwise coprime so that

$$
S_1 = \frac{N}{W} \tilde{\sum} \frac{\lambda_{d_1} \lambda_{e_1} \ldots \lambda_{d_k} \lambda_{e_k}}{\prod_{i=1}^{k} [d_i, e_i]} + O\left( \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k} |\lambda_{d_1} \lambda_{e_1} \ldots \lambda_{d_k} \lambda_{e_k}| \right). \quad (6.6)
$$

We need to bound our error term so we can check it is not too large. Define $\tau_r(n)$ as the number of ways of writing $n$ as a product of $r$ numbers; then, as with $y_{\text{max}}$ in the statement of Lemma 6.1, take $\lambda_{\text{max}} = \sup |\lambda_{d_1} \ldots \lambda_{d_k}|$ and recall $\prod_{i=1}^{k} d_i < R$, giving

$$
\sum_{d_1, \ldots, d_k, e_1, \ldots, e_k} |\lambda_{d_1} \lambda_{e_1} \ldots \lambda_{d_k} | \ll \lambda_{\text{max}}^2 \times (\text{number of terms in sum})
$$

$$
\ll \lambda_{\text{max}}^2 \left( \sum_{d \leq R} \tau_k(d) \right)^2. \quad (6.7)
$$

Now we have to bound this sum; luckily for us, this is quite standard by first using Rankin’s trick

$$
\sum_{d \leq R} \tau_k(d) \leq R \sum_{d \leq R} \frac{\tau_k(d)}{d},
$$
then splitting $d$ into its prime factors,

$$
= R \sum_{d_1, \ldots, d_k \leq R} \frac{\tau_k(\prod_{i=1}^k d_i)}{\prod_{d_i \leq R} d_i}.
$$

By definition $\tau_k(\prod_{i=1}^k d_i)$ is 1 (as the $d_i$ are squarefree and pairwise coprime), giving,

$$
= R \sum_{d_1, \ldots, d_k \leq R} \frac{1}{\prod_{d_i \leq R} d_i},
$$

and if we change $\prod d_i \leq R$ to just $d_i \leq R$ we get

$$
\leq R \left( \sum_{d \leq R} \frac{1}{d} \right)^k 
\ll R(\log R)^k.
$$

Putting it together we find

$$
\lambda_{\max}^2 \left( \sum_{d \leq R} \tau_k(d) \right)^2 \ll \lambda_{\max}^2 R^2 (\log R)^{2k},
$$

which will become negligible.

We need to have $\lambda_{d_1, \ldots, d_k}$ and $\lambda_{e_1, \ldots, e_k}$ independent so that shortly we can substitute them both out for the same term, however currently there is a condition on $[d_i, e_i]$ in our summation, which necessarily introduces a dependence. We shall remove this by using the fact that

$$
\frac{1}{[d_i, e_i]} = \frac{1}{d_i e_i} \sum_{u_i | d_i, e_i} \varphi(u_i).
$$
Letting $M$ be the main term of $S_1$ this gives

$$M = \frac{N}{W} \sum_{d_1,\ldots,d_k} \left( \lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k} \prod_{i=1}^{k} \left( \frac{1}{d_i e_i} \sum_{u_i | d_i, e_i} \varphi(u_i) \right) \right)$$

$$= \frac{N}{W} \sum_{d_1,\ldots,d_k} \left( \frac{\lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k}}{(\prod d_i) (\prod e_i)} \prod_{u_i | d_i, e_i} \varphi(u_i) \right)$$

$$= \frac{N}{W} \sum_{u_1,\ldots,u_k} \left( \prod_{i=1}^{k} \varphi(u_i) \right) \sum_{d_1,\ldots,d_k} \left( \lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k} \prod_{u_i | d_i, e_i} \varphi(u_i) \right).$$

(6.8)

The last equality requires pulling out the $\varphi(u_i)$, changing the order of operations and changing the indexing. That’s quite a lot for one step, so we’ll elucidate enough that you may believe it’s true. We’ll show how it works for the case $k = 2$ and note that the method extends in a natural way for any $k$.

First notice that the only thing that changes is that the $\prod (\sum \varphi(u_i))$ moves outside $\sum$ and the condition $u_i | d_i, e_i$ is added to the second sum, i.e. we may ignore (for the purpose of this exercise) the other summands. Take

$$\sigma = \sum_{d_1,d_2} \prod_{i=1}^{2} \sum_{u_i | d_i} \varphi(u_i).$$

This is a sum over all divisors $u_i$ of $d_i$, we’ll add an index to make this clear. Let $D_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,n_i}\}$ be the set of divisors of $d_i$, then a change to our new notation of $u_{i,k}$ (made in the first equality) gives

$$\sigma = \sum_{d_1,d_2} \prod_{i=1}^{2} \sum_{u_i \in D_i} \varphi(u_i) = \sum_{d_1,d_2} \prod_{i=1}^{2} \sum_{j=1}^{n_i} \varphi(u_{i,j})$$

$$= \sum_{d_1,d_2} \left( \sum_{j=1}^{n_i} \varphi(u_{1,j}) \right) \left( \sum_{k=1}^{n_i} \varphi(u_{2,k}) \right).$$

29
\[\sum_{d_1, d_2} \varphi(u_{1,1})\varphi(u_{2,1}) + \cdots + \varphi(u_{1,n_1})\varphi(u_{2,1}) + \cdots + \varphi(u_{1,n_1})\varphi(u_{2,n_2})\]

\[= \sum_{d_1, d_2} \sum_{1 \leq j \leq n_1} \varphi(u_{1,j})\varphi(u_{2,k})\]

\[= \sum_{d_1, d_2} \sum_{u_{i,j} \in D_i} \prod_{k=1}^2 \varphi(u_{k,j}).\]

So far so good; we’ve swapped the summation and the product by a simple expansion argument. The next thing to see is that we can either choose our \(d_1, d_2\) and then take \(u_i \mid d_i\), or choose our \(u_i\) first and then take the corresponding \(d_i\):

\[\sum_{d_1, d_2} \sum_{u_{i,j} \in D_i} \prod_{i=1}^2 \varphi(u_{i,j}) = \sum_{u_{1, u_2}, i=1}^2 \prod_{u_i \mid d_i} 1,\]

which is the formulation we are looking for. This extends in a natural way to \(d_1, \ldots, d_k\) to give the equality in (6.8).

The \(\sum\) notation has served us well, but it’s time to be rid of it in favour of a standard sum. Let’s see how much of a restriction it is.

In \(S_1\) we have:

- \(d_i\) is coprime to \(W\) for all \(i\) (so also \((e_i, W) = 1\))
- \((d_i, d_j) = 1\) for all \(i \neq j\),

so the only remaining part of \(\sum\) we’re missing is \((d_i, e_j) = 1\) whenever \(i \neq j\).

We’ll add a function to our summand that has that effect.

We know

\[\sum_{d \mid n} \mu(d) = \begin{cases} 
    1 & n = 1 \\
    0 & n \neq 1
\end{cases},\]

so

\[\prod_{1 \leq i, j \leq k} \left( \sum_{i \neq j} \mu(s_{i,j}) \right) \quad (6.9)\]

30
is equivalent to the additional condition (as if any \((d_i, e_j) \neq 1\) then the whole product becomes 0). This gives

\[
M = \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \varphi(u_i) \right) \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k} \left( \frac{\lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k}}{(\prod d_i) (\prod e_i)} \prod_{1 \leq i,j \leq k}^{i \neq j} \left( \sum_{s_{i,j}|(d_i, e_j)} \mu(s_{i,j}) \right) \right)
\]

(6.10)

and then, as with \(\varphi\), we can take this product outside to get

\[
S_1 = \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \varphi(u_i) \right) \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k} \left( \prod_{1 \leq i,j \leq k}^{i \neq j} \left( \sum_{s_{i,j}|(d_i, e_j)} \mu(s_{i,j}) \right) \right) \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k} \left( \frac{\lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k}}{(\prod d_i) (\prod e_i)} \right).
\]

(6.11)

Here we go now - the part you’ve all been waiting for - it’s time to introduce a change of variables. Let

\[
y_{r_1, \ldots, r_k} = \left( \prod_{i=1}^{k} \mu(r_i) \varphi(r_i) \right) \sum_{d_1, \ldots, d_k} \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} d_i}. \tag{6.12}
\]

Then any \(y_{r_1, \ldots, r_k}\) satisfying similar conditions to those we have on \(\lambda_{d_1, \ldots, d_k}\) (specifically: \(r = \prod_{i=1}^{k} r_i < R, (r, W) = 1\) and \(\mu(r)^2 = 1\)) corresponds to a \(\lambda_{d_1, \ldots, d_k}\) that fulfils our requirements because the change is invertible (which we will now show):

\[
\frac{y_{r_1, \ldots, r_k}}{\prod \varphi(r_i)} = \prod_{i=1}^{k} \mu(r_i) \sum_{d_1, \ldots, d_k} \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} d_i}
\]

\[
\sum_{r_1, \ldots, r_k} \frac{y_{r_1, \ldots, r_k}}{\prod \varphi(r_i)} = \sum_{r_1, \ldots, r_k} \prod_{i=1}^{k} \mu(r_i) \sum_{d_1, \ldots, d_k} \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} d_i}
\]

\[
= \sum_{r_1, \ldots, r_k} \prod_{i=1}^{k} \mu(r_i) \sum_{d_1, \ldots, d_k} \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} d_i}. \tag{6.13}
\]
We need to do a bit more manipulation using properties of the Möbius function to get this in an accessible form. As each \( e_i, r_i, d_i \) triple is independent, we can just show how this works with a single variable and it extends very easily to the product.

Taking

\[
\sum_{\substack{e_i \mid r_i \quad r_i \mid d_i}} \mu(r_i)
\]

set \( r_i = e_i r_i' \), then using that \( r_i \) is squarefree, we can split up \( \mu(r_i) \) into \( \mu(r_i') \mu(e_i) \) and the sum becomes

\[
= \sum_{\substack{r_i' \mid d_i \quad e_i \mid r_i}} \mu(r_i') \mu(e_i)
\]

\[
= \mu(e_i) \sum_{\substack{r_i' \mid d_i \quad e_i \mid r_i}} \mu(r_i').
\]

But this sum is zero if \( \frac{d_i}{e_i} \) is not 1, so

\[
= \begin{cases} 
\mu(e_i) & d_i = e_i \\
0 & d_i \neq e_i
\end{cases}
\]

which, in turn, thanks to our joint \( e_i \mid r_i \) and \( r_i \mid d_i \) conditions, gives that \( e_i = d_i \) whenever this sum is non-zero, so we get that (6.12) becomes

\[
\prod_{i=1}^{k} \mu(e_i) \sum_{e_1, \ldots, e_k} \frac{\lambda_{e_1, \ldots, e_k}}{\prod_{i=1}^{k} e_i}.
\]

The \( e_i \) have been prescribed by the \( d_i \) so we can also remove the sum over all \( e_i \) to get

\[
= \lambda_{e_1, \ldots, e_k} \prod_{i=1}^{k} \mu(e_i)
\]

But \( \mu(x)^2 = 1 \) whenever \( \mu(x) \) is not zero, so every time it contributes to the sum it is also a self inverse, that is, \( \mu(x) = \frac{1}{\mu(x)} \), which we substitute in mostly for convenience of the notation:

\[
\sum_{\substack{r_1, \ldots, r_k \quad e_i \mid r_i \quad r_i \mid \varphi(r_i)}} \frac{y_{r_1, \ldots, r_k}}{\prod_{i=1}^{k} \varphi(r_i)} = \frac{\lambda_{e_1, \ldots, e_k}}{\prod_{i=1}^{k} \mu(e_i) e_i}.
\]

(6.14)
We also find that
\[ \lambda_{\text{max}} \ll y_{\text{max}} (\log R)^k. \] (6.15)
Rearranging (6.12) and then substituting it and (6.15) into (6.11) (and again using that \( \mu(x) = \frac{1}{\mu(x)} \)) we get:

\[
S_1 = \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \varphi(u_i) \right) \sum_{1 \leq i, j \leq k} \left( \prod_{\substack{1 \leq i, j \leq k \atop i \neq j}} \mu(s_{i,j}) \right) \\
\times \prod_{i=1}^{k} \left( \frac{\mu(a_i)\mu(b_i)}{\varphi(a_i)\varphi(b_i)} \right) y_{a_1, a_2, \ldots, a_k} y_{b_1, \ldots, b_k} + O \left( y_{\text{max}}^2 R^2 (\log R)^{4k} \right),
\] (6.16)

where \( a_i = u_i \prod_{i \neq j} s_{i,j} \) and \( b_j = u_j \prod_{i \neq j} s_{i,j} \). We note \( a_i \) is squarefree whenever the summand is non-zero, because \( (u_i, s_{i,j}) \neq 1 \) implies \( (e_i, e_j) \neq 1 \) and if for some \( \eta \neq \zeta, (s_{i,\eta}, s_{i,\zeta}) \neq 1 \) then we also get \( (e_\eta, e_\zeta) \neq 1 \), in either case our \( \lambda_{e_1, \ldots, e_k} \) are not supported on such \( e_i \). In the same way \( b_j \) is squarefree. Then \( \mu(a_i) = \mu(u_i) \prod_{i \neq j} \mu(s_{i,j}) \), and we can also rewrite \( \mu(b_j), \varphi(a_i) \) and \( \varphi(b_j) \) in a similar way to give

\[
\prod_{i=1}^{k} \left( \frac{\mu(a_i)\mu(b_i)}{\varphi(a_i)\varphi(b_i)} \right) = \prod_{i=1}^{k} \left( \frac{\mu(u_i)^2}{\varphi(u_i)^2} \right) \prod_{i \neq j} \left( \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2} \right),
\]

which changes our main sum to

\[
S_1 = \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \varphi(u_i) \right) \sum_{1 \leq i, j \leq k} \left( \prod_{\substack{1 \leq i, j \leq k \atop i \neq j}} \mu(s_{i,j}) \right) \\
\times \prod_{i=1}^{k} \left( \frac{\mu(u_i)^2}{\varphi(u_i)^2} \right) y_{a_1, a_2, \ldots, a_k} y_{b_1, \ldots, b_k} + O \left( y_{\text{max}}^2 R^2 (\log R)^{4k} \right)
\]

\[
= \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \varphi(u_i) \right) \sum_{1 \leq i, j \leq k} \left( \prod_{\substack{1 \leq i, j \leq k \atop i \neq j}} \frac{\mu(s_{i,j})^3}{\varphi(s_{i,j})^2} \right) \\
\times \prod_{i=1}^{k} \left( \frac{\mu(u_i)^2}{\varphi(u_i)^2} \right) y_{a_1, a_2, \ldots, a_k} y_{b_1, \ldots, b_k} + O \left( y_{\text{max}}^2 R^2 (\log R)^{4k} \right).
\]
By the nature of $\mu, \mu(x)^3 = \mu(x)$ for any $x$. So we have

$$= \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \frac{\mu(u_i)^2}{\varphi(u_i)} \right) \sum_{1 \leq i, j \leq k} \prod_{\substack{i \leq j \leq k \atop i \neq j}} \frac{\mu(s_{i,j})}{\varphi(s_{i,j})^2} \right) \right) \tag{6.17}$$

$$\times y_{a_1, \ldots, a_k} y_{b_1, \ldots, b_k} + O \left( y_{\max}^2 R^2 (\log R)^{4k} \right).$$

Now we split the main term into the cases $s_{i,j} = 1$ and $s_{i,j} \neq 1$.

Taking the second case first, notice as $s_{i,j} | (d_i, e_j)$ and $(d_i, W) = (e_j, W) = 1$ then also $(s_{i,j}, W) = 1$, so we must have $s_{i,j} > D_0$. Then the contribution is

$$\ll y_{\max}^2 \frac{N}{W} \left( \sum_{u \leq R} \frac{\mu(u)^2}{\varphi(u)} \right) \prod_{\substack{i \neq j \atop 1 \leq i, j \leq k}} \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2}. \right)$$

Then, as $\frac{\mu(a)}{\varphi(a)} \leq 1$ for all $a$, we have

$$\ll y_{\max}^2 \frac{N}{W} \prod_{i=1}^{k} \left( \sum_{u \leq R} \frac{\mu(u)^2}{\varphi(u)} \right) \prod_{i \leq j \neq 1} \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2}.$$ 

For both products, we remove the dependence on $i$ and $j$, so that each term in the product is identical, giving

$$\ll y_{\max}^2 \frac{N}{W} \left( \sum_{u \leq R} \frac{\mu(u)^2}{\varphi(u)} \right)^k \left( \sum_{s_{i,j} > D_0} \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2} \right)^{k^2-k}. \right)$$

Then, as before, because $\frac{\mu(a)}{\varphi(a)} \leq 1$ for all $a$, we have

$$\sum_{s_{i,j} > D_0} \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2} \leq \sum_{s \geq 1} \frac{\mu(s)^2}{\varphi(s)^2}.$$
We can take \( k^2 - k - 1 \) terms out of our second sum and apply this inequality to give:

\[
\ll y_{\text{max}}^2 \frac{N}{W} \left( \sum_{u < R} \mu(u)^2 \phi(u) \right)^k \left( \sum_{s_{i,j} > D_0} \mu(s_{i,j})^2 \phi(s_{i,j})^2 \right) \left( \sum_{s \geq 1} \mu(s)^2 \phi(s)^2 \right)^{k^2 - k - 1}.
\]

(6.18)

For the result we need to bound each part in this sum. We’re going to work from the last factor to the first.

We recall that \( \phi(s) \approx s \), particularly for large \( s \). So heuristically we get

\[
\sum_{s \geq 1} \frac{\mu(s)^2}{\phi(s)^2} \approx \sum_{s \geq 1} \frac{1}{s^2} = \frac{\pi^2}{6}.
\]

As the right hand side converges to a constant, so does the left.

Next, we’re taking exactly the same sum, but starting a little further along:

\[
\sum_{s_{i,j} > D_0} \frac{\mu(s_{i,j})^2}{\phi(s_{i,j})^2} \approx \sum_{s \geq D_0} \frac{1}{s^2} = \sum_{s \geq 1} \frac{1}{s^2} - \sum_{s = 1}^{D_0} \frac{1}{s^2} = \frac{\pi^2}{6} - \frac{1}{D_0} = \frac{C}{D_0},
\]

for a constant \( C \) which will be absorbed into our \( \ll \) constant.

Finally, we have just one term left:

\[
\sum_{u < R} \frac{\mu(u)^2}{\phi(u)}.
\]

However we’re going to need an additional lemma to bound this:
Lemma 6.2. Let $U_1, U_2, L > 0$. Let $\gamma$ be a multiplicative function satisfying

$$0 \leq \frac{\gamma(p)}{p} \leq 1 - U_1,$$

and

$$-L \leq \sum_{w \leq p \leq z} \frac{\gamma(p) \log p}{p} - \log \frac{z}{w} \leq U_2$$

for any $2 \leq w \leq z$. Let $\eta$ be the totally multiplicative function defined on primes by

$$\eta(p) = \frac{\gamma(p)}{p - \gamma(p)}.$$

Finally, let $G: [0, 1] \to \mathbb{R}$ be smooth, and let $G_{\text{max}} = \sup_{t \in [0,1]} (|G(t)| + |G'(t)|)$. Then

$$\sum_{d < z} \mu(d)^2 \eta(d) G \left( \frac{\log d}{\log z} \right) = P \log z \int_0^1 G(x) \, dx + O_{U_1, U_2}(PLG_{\text{max}})$$

where

$$P = \prod_{p \text{ prime}} \left( 1 - \frac{\gamma(p)}{p} \right)^{-1} \left( 1 - \frac{1}{p} \right).$$

Proof. As we’re only going to use this as a tool, the proof is not very enlightening. However, it can be found in [5, Lemmas 3 & 4].

Getting back to bounding, we rewrite

$$\sum_{u < R \atop (u, W) = 1} \frac{\mu(u)^2}{\varphi(u)},$$

so that the lemma is clearly applicable. Take

$$\gamma(p) = \begin{cases} 1 & p \nmid W \\ 0 & p \mid W \end{cases}$$

and

$$G(x) = 1.$$
Then
\[
\sum_{d<z} \mu(d)^2 \eta(d) G \left( \frac{\log d}{\log z} \right) = \sum_{d<z} \mu(d)^2 \prod_{p|d} \frac{\gamma(p)}{p - \gamma(p)}
\]
\[
= \sum_{d<z} \mu(d)^2 \prod_{p|d} (p - 1)
\]
\[
= \sum_{d<R} \frac{\mu(d)^2}{\varphi(d)},
\]
(6.19)

which is exactly as we'd like. So to get our bound we just have to evaluate \( P \) for this \( \gamma \) (as the integral is trivially 1). We have
\[
\left( 1 - \frac{\gamma(p)}{p} \right) = \begin{cases} 
    \left( 1 - \frac{1}{p} \right) & p \nmid W \\
    1 & p | W,
\end{cases}
\]
so
\[
P = \prod_{p|W} \left( 1 - \frac{1}{p} \right)
\]
\[
= \prod_{p|W} (p - 1) \left( \prod_{p|W} p \right)^{-1}
\]
\[
= \frac{\varphi(W)}{W},
\]
(6.20)
giving us the bound
\[
\left( \sum_{u<R} \frac{\mu(u)^2}{\varphi(u)} \right) = \frac{\varphi(W)}{W} \log(R) + O \left( \frac{\varphi(W)}{W} \right)
\]
(6.21)
\[
= \frac{\varphi(W)}{W} (\log(R) + O(1)).
\]
but as this is an error term itself, the \( O(1) \) will disappear in the final error.
In total we now have contribution

$$
\ll y^2_{\max} \frac{N}{W} \left( \sum_{u < R} \frac{\mu(u)^2}{\varphi(u)} \right)^k \left( \sum_{s_{i,j} > D_0} \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2} \right) \left( \sum_{s \geq 1} \frac{\mu(s)^2}{\varphi(s)^2} \right)^{k^2 - k - 1}
$$

$$
= y^2_{\max} \frac{N}{W} \left( \frac{\varphi(W)}{W} (\log(R) + O(1)) \right)^k \left( \frac{C}{D_0} \right)
\ll y^2_{\max} N \varphi(W)^k (\log R)^k
\frac{k^2}{W^{k+1}D_0}
$$

from $s_{i,j} > D_0$.

The case $s_{i,j} = 1$ whenever $i \neq j$ is now our main term and we add on the error incurred by the $s_{i,j} > D_0$ to give us:

$$
S_1 = \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \frac{y^2_{u_1, \ldots, u_k}}{\prod_{i=1}^k \varphi(u_i)} \right) + O \left( \frac{y^2_{\max} \varphi(W)^k N (\log R)^k}{W^{k+1}D_0} + y^2_{\max} R^2 (\log R)^{4k} \right)
$$

of which the first error term dominates, giving the result

$$
S_1 = \frac{N}{W} \sum_{u_1, \ldots, u_k} \left( \frac{y^2_{u_1, \ldots, u_k}}{\prod_{i=1}^k \varphi(u_i)} \right) + O \left( \frac{y^2_{\max} \varphi(W)^k N (\log R)^k}{W^{k+1}D_0} \right).
$$

\hfill \square

6.3 $S_2$

Phew! That’s half the sum done. We will follow much the same overall scheme for $S_2$, but there are some non-trivial technical differences between the two sums, even though the two lemmas may look very similar.

We have

$$
S_2 = \sum_{N \leq n \leq 2N \atop n \equiv v(W)} P_n w_n,
$$

which we split up into

$$
S_2 = \sum_{m=1}^k S_2^{(m)},
$$

38
where

\[ S_2^{(m)} = \sum_{N \leq n \leq 2N} \chi_P(n + h_m) \left( \sum_{d_1,\ldots,d_k \mid n+h_i} \lambda_{d_1,\ldots,d_k} \right)^2. \]

Now we have the equivalent of Lemma 6.1 for \( S_2^{(m)} \).

Define \( g(x) \) to be the multiplicative function defined on the primes by \( g(p) = p - 2 \) and analogously to (6.2), define

\[ y_{\text{max}}^{(m)} = \sup_{r_1,\ldots,r_k} \left| y_r^{(m)} \right|. \]

Lemma 6.3. Let

\[ y_r^{(m)} = \left( \prod_{i=1}^k \mu(r_i) g(r_i) \right) \sum_{\substack{d_1,\ldots,d_k \mid r_i \mid d_i \mid d_m = 1 \n \equiv v(W) \mid d_i, e_i \mid n \equiv v(W) \mid [d_i, e_i] \mid n+h_i}} \frac{\lambda_d}{\prod_{i=1}^k \varphi(d_i)}. \tag{6.22} \]

Then for any \( A > 0 \)

\[ S_2^{(m)} = \frac{N}{\varphi(W) \log N} \sum_{r_1,\ldots,r_k} \frac{(y_r^{(m)})^2}{\prod_{i=1}^k g(r_i)} + O \left( \frac{(y_{\text{max}}^{(m)})^2 \varphi(W)^{k-2} N (\log R)^{k-2}}{W^{k-1} D_0} \right) \]

\[ + O \left( \frac{y_{\text{max}}^{(m)}}{(\log N)^A} \right). \tag{6.23} \]

Proof. Recall that we use \([a, b]\) to refer to the lowest common multiple of \( a \) and \( b \).

In the same way as with \( S_1 \), we change the order of summation to put the \( \lambda_d \) in a sum just depending on \( d_1,\ldots,d_k \) and \( e_1,\ldots,e_k \):

\[ S_2^{(m)} = \sum_{\substack{d_1,\ldots,d_k \mid n+h_i \n \equiv v(W) \mid [d_i, e_i] \mid n \equiv v(W) \mid [d_i, e_i] \mid n+h_i}} \lambda_{d_1,\ldots,d_k} \chi_P(n + h_m). \]

Again, we look at the inner sum first, although this time its expansion is not quite so obvious (we’ll explain it after setting it all out).

Define

\[ q := W \prod_{i=1}^k [d_i, e_i], \]

39
Then
\[
\sum_{N \leq n \leq 2N} \chi_P(n + h_m)
\]
\[
= \sum_{N \leq n \leq 2N} \frac{\chi_p(n)}{\varphi(q)} + O \left( 1 + \sup_{(a, q) = 1} \left| \sum_{N \leq n \leq 2N \atop n \equiv a (q)} \chi_p(n) - \frac{1}{\varphi(q)} \sum_{n \leq 2N} \chi_p(n) \right| \right).
\]

For ease of notation, we’ll set
\[
X_N = \sum_{N \leq n \leq 2N} \chi_p(n)
\]
and
\[
E(N, q) = O \left( 1 + \sup_{(a, q) = 1} \left| \sum_{N \leq n \leq 2N \atop n \equiv a (q)} \chi_p(n) - \frac{1}{\varphi(q)} \sum_{n \leq 2N} \chi_p(n) \right| \right). \quad (6.24)
\]

Let’s look at each part of this separately.

We’re looking for the number of elements in \(\{n + h_m\}_{n=N}^{2N}\) that are prime, congruent to \(v\) modulo \(W\), and coprime to \(\prod [d_i, e_i]n + h_i\). We can immediately remove the \(h_m\) from the \(\chi_p\) function incurring an error of at most 1 which we included in the \(E(N, q)\) term. The number of elements that satisfy our congruence conditions is \(\frac{N}{\varphi(q)} + O(1)\), where, in this case, \(O(1)\) actually means \(\pm 1\) or 0. The number of those which we expect to be prime is then
\[
\frac{1}{\varphi(q)} \sum_{N \leq n \leq 2N} \chi_p(n) + O(1),
\]
but unfortunately this is not sufficient when dealing with primes. The prime numbers may not be evenly distributed among the congruence classes of \(q\), and this means we have to add an extra error term, the second part of our \(E(N, q)\) definition. This part says we’ll be wrong by at most the largest difference between what we expect to be true and what is actually true.
Now using the $\tilde{\sum}$ notation as before (summing only over cases when $W, [d_1, e_1], \ldots, [d_k, e_k]$ are all pairwise coprime) we have

$$S_2^{(m)} = \frac{X_n}{\varphi(W)} \sum_{d_1, \ldots, d_k \atop e_1, \ldots, e_k} \frac{\lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k}}{\varphi(\prod_{i=1}^k [d_i, e_i])} + O \left( \sum_{d_1, \ldots, d_k \atop e_1, \ldots, e_k} \lambda_{d_1, \ldots, d_k}^2 \lambda_{e_1, \ldots, e_k}^2 \lambda_{\text{max}} E(N, q) \right).$$

Following our treatment of $S_1$, we'll bound the error first, using the same bound of $\lambda_{\text{max}} \ll y_{\text{max}} (\log R)^k$. A cautious reader may worry that the slight (‘slight’ because if $n$ has no small factors then $\varphi(n) \approx n \approx \prod_{p|n} p - 2 = g(n)$) change in our definition of $y$ with $g(r_i)$ in place of $\varphi(r_i)$ will change this bound, but that change is held in the $y_{\text{max}}$ part of the bound. The other thing we want to bound is

$$\sum_{d_1, \ldots, d_k \atop e_1, \ldots, e_k} E(N, q) = \sum_{d_1, \ldots, d_k \atop e_1, \ldots, e_k} E(N, W \prod_{i=1}^k [d_i, e_i])$$

$$= \sum_{q} E(N, q) \left\{ \# \text{ of different } d_i, e_i \text{ with } \prod_{i=1}^k [d_i, e_i] = \frac{q}{W} \right\}.$$

As $\lambda_d$ are only non-zero when $\prod d_i \leq R$, we have that $q \leq R^2 W$. The $d_i$ were defined to be squarefree and $W$ clearly is by its definition, so we only need to consider squarefree $q$ in this range.

**Lemma 6.4.** For any square-free integer $n$, there are $\tau_{3k}(n)$ choices for $d_1, \ldots, d_k, e_1, \ldots, e_k$ such that $W \prod_{i=1}^k [d_i, e_i] = n$.

**Proof.** As $n$ is square-free, given $p | n$, we see that if $p \mid W$, then $p \nmid d_i$ and $p \nmid e_i$ for any $i$.

If instead, $p \nmid n$ and $p \nmid W$ then we must choose one of $k$ pairs $d_i, e_i$ such that $p \mid [d_i, e_i]$. Then we choose one of 3 options:

- $p \mid d_i \quad p \nmid e_i$,
- $p \nmid d_i \quad p \mid e_i$,
- $p \mid d_i \quad p \mid e_i$.  

41
In total this gives us $3k$ choices of which of $d_1, \ldots, d_k, e_1, \ldots, e_k$ is a multiple of $p$. Therefore there are at most

$$\prod_{p|n} 3k = \tau_{3k}(n)$$

choices of $d_1, \ldots, d_k, e_1, \ldots, e_k$ for any given $n$ (there will in fact be fewer choices whenever $W \neq 1$).

As in the manipulation of $S_1$, our first order of business is removing the dependence between $d_i$ and $e_j$. We do this by substituting out the $\varphi[d_i, e_i]$ term, but this time we use an identity for $g(x)$:

$$\frac{1}{\varphi[d_i, e_i]} = \frac{1}{\varphi(d_i)\varphi(e_i)} \sum_{u_i|d_i, e_i} g(u_i),$$

where $g$ is the totally multiplicative function defined by $g(p) = p - 2$ giving main term

$$M' = \frac{X_n}{\varphi(W)} \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k \atop d_m = e_m = 1} \lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k} \prod_{d_i \atop \varphi(d_i)} \left( \sum_{u_i|d_i, e_i} g(u_i) \right) / \prod_{d_i \atop \varphi(d_i)} \prod_{e_i} \varphi(e_i).$$

Again we want to remove the $\sum$ notation, so we include the factor of

$$\prod_{1 \leq i, j \leq k \atop i \neq j} \left( \sum_{s_{i,j}|d_i, e_j} \mu(s_{i,j}) \right)$$

and repeat our trick for pulling products out of the summation over $d_1, \ldots, d_k$.  

42
\[ e_1, \ldots, e_k \] to give us

\[ M' = \frac{X_n}{\varphi(W)} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} g(u_i) \right) \sum_{1 \leq i, j \leq k, i \neq j} \left( \prod_{1 \leq i, j \leq k, i \neq j} \mu(s_{i,j}) \right) \]

\[ \times \sum_{d_1, \ldots, d_k, e_1, \ldots, e_k} \left( \frac{\lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k}}{(\prod \varphi(d_i))(\prod \varphi(e_i))} \right). \]

If you’ve been comparing to Section 6.1 you’ll know what’s coming next, with just a slight adjustment to cater for the \( g(u_i) \) in place of \( \varphi(u_i) \). We make the substitution

\[ y_{r_1, \ldots, r_k}^{(m)} = \left( \prod_{i=1}^{k} \mu(r_i) g(r_i) \right) \sum_{d_1, \ldots, d_k, r_i | d_i \neq d_m = 1} \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} \varphi(d_i)} \]

which gives us

\[ M' = \frac{X_n}{\varphi(W)} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} g(u_i) \right) \sum_{1 \leq i, j \leq k, i \neq j} \left( \prod_{1 \leq i, j \leq k, i \neq j} \mu(s_{i,j}) \right) \]

\[ \times \prod_{i=1}^{k} \left( \frac{\mu(a_i) \mu(b_i)}{g(a_i)g(b_i)} \right) y_{a_1, \ldots, a_k}^{(m)} y_{b_1, \ldots, b_k}^{(m)} \]

\[ = \frac{X_n}{\varphi(W)} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \mu(u_i) g(u_i) \right) \sum_{1 \leq i, j \leq k, i \neq j} \left( \prod_{1 \leq i, j \leq k, i \neq j} \mu(s_{i,j}) \right) \frac{1}{g(s_{i,j})^2} y_{a_1, \ldots, a_k}^{(m)} y_{b_1, \ldots, b_k}^{(m)} \]

\[ = \frac{X_n}{\varphi(W)} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} \mu(u_i) g(u_i) \right) \sum_{1 \leq i, j \leq k, i \neq j} \left( \prod_{1 \leq i, j \leq k, i \neq j} \mu(s_{i,j}) \right) \frac{1}{g(s_{i,j})^2} y_{a_1, \ldots, a_k}^{(m)} y_{b_1, \ldots, b_k}^{(m)}, \]

with \( a_i = u_i \prod_{i \neq j} s_{i,j} \) and \( b_j = u_j \prod_{i \neq j} s_{i,j} \) as before. Then we split into the cases for \( s_{i,j} > D_0 \) and \( s_{i,j} = 1 \). We follow an identical method to get an
analogous result to (6.18), this time with \( g(u) \) in place of \( \varphi(u) \):

\[
(y_{\text{max}}^{(m)})^2 \frac{X_n}{\varphi(W)} \left( \sum_{u < R} \frac{\mu(u)^2}{g(u)} \right)^{k-1} \left( \sum_{s_i, j > D_0} \frac{\mu(s_i, j)^2}{g(s_i, j)^2} \right) \left( \sum_{s \geq 1} \frac{\mu(s)^2}{g(s)^2} \right)^{k^2 - k - 1}.
\]

(6.26)

Then we bound the sums as before, along with estimating \( X_n \) to be \( \frac{N}{\log N} \), to get

\[
\ll (y_{\text{max}}^{(m)})^2 \frac{N}{\varphi(W) \log N} \left( \frac{\varphi(W) \log R}{W} \right)^{k-1} \frac{1}{D_0} = (y_{\text{max}}^{(m)})^2 \frac{\varphi(W)^{k-2} N (\log R)^{k-1}}{W^{k-1} D \log N}.
\]

Thus we get the result that

\[
S_2^{(m)} = \frac{X_n}{\varphi(W)} \sum_{u_1, \ldots, u_k} \left( \prod_{i=1}^{k} g(u_i) \right) y_{u_1, \ldots, u_k} + O \left( (y_{\text{max}}^{(m)})^2 \frac{\varphi(W)^{k-2} N (\log R)^{k-1}}{W^{k-1} D \log N} \right).
\]

\[
\square
\]

7 The Lagrangian Method

Our aim is to make the value of \( S_1 \) as small as possible while making \( S_2 \) large. This idea of minimising with a constraint suggests using the method of Lagrangian multipliers, which is exactly what we do now.

Taking just the main terms, we have

\[
S_1 = \frac{N}{W} \sum_{r_1, \ldots, r_k} y_{r_1, \ldots, r_k}^2 \prod_{i=1}^{k} \varphi(r_i),
\]

\[
S_2 = \sum_{m=1}^{k} \frac{N}{\varphi(W) \log N} \sum_{r_1, \ldots, r_k} \frac{(y_{z}^{(m)})^2}{\prod_{i=1}^{k} g(r_i)}.
\]

There are two tricks we can use right from the start: first we can drop all constants because it won’t affect which choices of \( y_z \) are minimal; second, we
note that scaling $S_1$ scales $S_2$ an equivalent amount so we can set
\[ \sum_{r_1, \ldots, r_k} \frac{y_{r_1, \ldots, r_k}^2}{\prod_{i=1}^k \varphi(r_i)} = 1, \]
and this will be our constraint for the Lagrangian multipliers method.

\[
L(y, \lambda) = S_2 - \lambda (S_1 - 1) \\
L(y, \lambda) = \sum_{m=1}^k \sum_{r_1, \ldots, r_k} \frac{(y_{r_1, \ldots, r_k}^{(m)})^2}{\prod g(r_i)} - \lambda \left( \sum_{r_1, \ldots, r_k} \frac{y_{r_1, \ldots, r_k}^2}{\prod \varphi(r_i)} - 1 \right). \tag{7.1}
\]

So the system of equations we are trying to solve is:
\[
\frac{\partial L}{\partial y_{u_1, \ldots, u_k}} = 0 \\
\frac{\partial L}{\partial y_{r_1, \ldots, r_k}} = 0.
\]

For both of these equations we have to be clear on how $y_{r_1, \ldots, r_k}$ depends on $y_{u_1, \ldots, u_k}$ and so solving the second one gets us a long way towards solving the first. Recall the definitions of $y_{r_1, \ldots, r_k}$ and $y_{u_1, \ldots, u_k}$:
\[
y_{r_1, \ldots, r_k} = \left( \prod_{i=1}^k \mu(r_i) g(r_i) \right) \sum_{d_1, \ldots, d_k} \frac{\lambda_d}{\prod_{i=1}^k \varphi(d_i)} \\
y_{u_1, \ldots, u_k} = \left( \prod_{i=1}^k \mu(u_i) \varphi(u_i) \right) \sum_{d_1, \ldots, d_k} \frac{\lambda_d}{\prod_{i=1}^k \varphi(d_i)}.
\]

To evaluate $\frac{\partial y_{r_1, \ldots, r_k}}{\partial y_{u_1, \ldots, u_k}}$, it is necessary for us to have an expression for $y_{r_1, \ldots, r_k}$ in terms of $y_{u_1, \ldots, u_k}$; we shall get this with the following lemma.

**Lemma 7.1.** If $r_m = 1$ then
\[
y_{r_1, \ldots, r_k} = \sum_{a_m} \frac{y_{r_1, \ldots, r_{m-1}, a_m, r_{m+1}, \ldots, r_k}}{\varphi(a_m)} + O \left( \frac{y_{\text{max}} \varphi(W) \log R}{WD_0} \right). \tag{7.2}
\]
Proof. Assume \( r_m = 1 \). Then, from (6.14), we have

\[
\sum_{r_1, \ldots, r_k} \frac{y_{r_1, \ldots, r_k}}{d_1 | r_1} = \frac{\lambda_{d_1, \ldots, d_k}}{\prod_{i=1}^{k} \mu(d_i) d_i}
\]

and from (6.22),

\[
y^{(m)}_\mu = \left( \prod_{i=1}^{k} \mu(r_i) g(r_i) \right) \sum_{d_1, \ldots, d_k} \prod_{i=1}^{k} \frac{\lambda_i}{\mu(d_i) d_i} \prod_{r_i | d_i} \varphi(d_i). \tag{7.3}
\]

A moment of care must be taken here to clarify that the vectors \( r \) are dummy variables and not the same in the two definitions, henceforth we change the \( y_r \) to \( y_a \) to avoid confusion. Now substituting for \( \lambda_d \) we get

\[
y^{(m)}_\mu = \left( \prod_{i=1}^{k} \mu(r_i) g(r_i) \right) \sum_{d_1, \ldots, d_k} \prod_{i=1}^{k} \frac{\lambda_i}{\mu(d_i) d_i} \sum_{a_1, \ldots, a_k} \prod_{i=1}^{k} \frac{y_a}{\varphi(d_i)}. \tag{7.3}
\]

We want to swap the order of the summations to get them into forms we can evaluate. Currently, we choose \( r_i \) first, then \( d_i \) and finally choose the \( a_i \) such that \( d_i | a_i \); instead, we want to choose the \( a_i \) before the \( d_i \) while keeping all other conditions the same.

As we have \( r_i | d_i \) and \( d_i | a_i \), we have \( r_i | a_i \) and having selected the \( a_i \) we can then keep the \( d_i | a_i \) condition by imposing it on the second sum, giving us

\[
y^{(m)}_\mu = \left( \prod_{i=1}^{k} \mu(r_i) g(r_i) \right) \sum_{a_1, \ldots, a_k} \prod_{r_i | a_i} \frac{y_a}{\varphi(a_i)} \sum_{d_1, \ldots, d_k} \prod_{r_i | d_i} \prod_{d_i | a_i} \frac{\mu(d_i) d_i}{\varphi(d_i)}. \tag{7.4}
\]

By the multiplicative nature of \( \mu \) and \( \varphi \), this is

\[
y^{(m)}_\mu = \left( \prod_{i=1}^{k} \mu(r_i) g(r_i) \right) \sum_{a_1, \ldots, a_k} \frac{y_a}{\prod_{r_i | a_i} \varphi(a_i)} \prod_{i \neq m} \frac{\mu(a_i) r_i}{\varphi(a_i)}. \tag{7.4}
\]

We note that \( y_a \) is 0 whenever \( (a_j, W) \neq 1 \) so either \( a_j = r_j \) or \( a_j > D_0 r_j \).
As for (6.17), we will take $a_j > D_0r_j$ first and show it has negligible impact. The sum (7.4) with the caveat $a_j > D_0r_j$ is less than a constant times:

$$\left( \prod_{i=1}^{k} \mu(r_i)g(r_i) \right) \sum_{a_1,\ldots,a_k \atop r_i \mid a_i} \frac{y_{\text{max}}}{\prod_{i=1}^{k} \varphi(a_i)} \prod_{i \neq j} \frac{\mu(a_i)r_i}{\varphi(a_i)}. \quad (7.5)$$

Previously, we chose $r_i$ first, then subsequently the corresponding $a_i$, and then take a product over all the $\varphi(a_i)$. This means that when we make the $a_i$ conditions separate we have to square the product to compensate for this double counting. We do this, as well as splitting the sum into $i = m$ and $i \neq m$ summands, to get:

$$\ll y_{\text{max}} \left( \prod_{i=1}^{k} g(r_i)r_i \right) \left( \sum_{a_1,\ldots,a_k} \frac{\mu(a_m)^2}{\varphi(a_m)} \right) \left( \sum_{a_1,\ldots,a_k} \frac{\mu(a_i)^2}{\varphi(a_i)^2} \right) \sum_{a_{i} > D_0} \prod_{i \neq j \atop i \neq m} \frac{\mu(a_i)^2}{\varphi(a_i)^2},$$

$$\ll y_{\text{max}} \left( \prod_{i=1}^{k} g(r_i)r_i \right) \left( \sum_{a_{i} < R} \frac{\mu(a_m)^2}{(a_m,W)^2} \right) \left( \sum_{a_{i} > D_0} \frac{\mu(a_i)^2}{\varphi(a_i)^2} \right) \prod_{i \neq j \atop i \neq m} \frac{\mu(a_i)^2}{\varphi(a_i)^2},$$

each part of which we can evaluate as we did for (6.17), leaving us with

$$\ll y_{\text{max}} \left( \prod_{i=1}^{k} g(r_i)r_i \right) \frac{\varphi(W) \log R}{WD_0},$$

$$\ll y_{\text{max}} \frac{\varphi(W) \log R}{WD_0}. \quad (7.6)$$

47
This gives us that the main term is when \( a_j = r_j \) for all \( j \neq m \) and so

\[
y^{(m)}_x = \left( \prod_{i=1}^{k} \frac{\mu(r_i) g(r_i)}{\varphi(r_i)} \right) \sum_{a_m} \left( \frac{y_{r_1, \ldots, r_{m-1}, a_m, r_{m+1}, \ldots, r_k}}{\prod_{i=1}^{k} \varphi(a_i)} \right) \prod_{i \neq m} \frac{\mu(a_i) r_i}{\varphi(a_i)} + O \left( \frac{y_{\text{max}} \varphi(W) \log R}{W D_0} \right)
\]

\[
= \left( \prod_{i=1}^{k} \frac{\mu(r_i) g(r_i)}{\varphi(r_i)} \right) \sum_{a_m} \left( \frac{y_{r_1, \ldots, r_{m-1}, a_m, r_{m+1}, \ldots, r_k}}{\varphi(a_m)} \right) \prod_{i \neq m} \frac{\mu(r_i) r_i}{\varphi(r_i)} + O \left( \frac{y_{\text{max}} \varphi(W) \log R}{W D_0} \right)
\]

\[
= \left( \prod_{i=1}^{k} \frac{\mu(r_i)^2 g(r_i) r_i}{\varphi(r_i)^2} \right) \sum_{a_m} \left( \frac{y_{r_1, \ldots, r_{m-1}, a_m, r_{m+1}, \ldots, r_k}}{\varphi(a_m)^2} \right) \frac{\varphi(a_m)^2}{\mu(a_m) a_m} + O \left( \frac{y_{\text{max}} \varphi(W) \log R}{W D_0} \right)
\]

Finally we note that

\[
\frac{g(x) x}{\varphi(x)^2} = 1 + O(x^{-2}).
\]

As our \( y^{(m)}_x \) are only supported when \( \prod_{i=1}^{k} r_i > D_0 \), this can be compensated for with a

\[
1 + O(D_0^{-2})
\]

term. This is still less than our other error term. Finally, we see that both \( y^{(m)}_x \) and \( y^{(m)}_u \) are zero when the \( r_i \) are not square free, so we can drop the \( \mu(r_i)^2 \) term to get

\[
y^{(m)}_{r_1, \ldots, r_k} = \sum_{a_m} \frac{y_{r_1, \ldots, r_{m-1}, a_m, r_{m+1}, \ldots, r_k}}{\varphi(a_m)} + O \left( \frac{y_{\text{max}} \varphi(W) \log R}{W D_0} \right).
\]

As a result of this lemma, we have that \( y^{(m)}_x \) is a sum of terms almost all of which do not depend on \( y_u \). For dependence on \( y_u \) we need \( r_i = u_i \) for all \( i \neq m \); in which case, the previous lemma gives, for \( r_m = 1 \),

\[
y^{(m)}_x = \sum_{a_m} \frac{y_{r_1, \ldots, r_{m-1}, a_m, 1, r_{m+1}, \ldots, r_k}}{\varphi(a_m)} + O \left( \frac{y_{\text{max}} \varphi(W) \log R}{W D_0} \right).
\]  

\[\text{(7.7)}\]
Hence,
\[
\frac{\partial y^{(m)}_L}{\partial y_u} = \begin{cases} 
\sum_{a_m} \frac{1}{\varphi(a_m)}, & \text{if } r_i = u_i \text{ for all } i \neq m \\
0, & \text{otherwise}
\end{cases}.
\] (7.8)

Which gives in total for \( L \)
\[
\frac{\partial L}{\partial y_{u_1,\ldots,u_k}} = \frac{\partial}{\partial y_u} \sum_{m=1}^{k} \sum_{r_1,\ldots,r_k} \left( y^{(m)}_L \right)^2 \prod_{i \neq m} g(r_i) - \lambda \frac{\partial}{\partial y_u} \left( \sum_{r_1,\ldots,r_k} \prod_{i \neq m} \varphi(r_i) \right)
\]
\[
= \sum_{m=1}^{k} \sum_{r_1,\ldots,r_k} 2 \frac{\partial (y^{(m)}_L)}{\partial y_u} \frac{y^{(m)}_L}{\prod_{i \neq m} g(r_i)} - \lambda \left( \sum_{r_1,\ldots,r_k} 2 \frac{\partial y_L}{\partial y_u} \prod_{i \neq m} \varphi(r_i) \right).
\]

The sum over the \( r_i \) has all terms 0 except for when the \( r_i = u_i, i \neq m \), so we can remove this sum and instead evaluate
\[
\sum_{m=1}^{k} \prod_{i \neq m} g(u_i) \frac{1}{\varphi(a_m)} \sum_{a_m} 2 y^{(m)}_{u_1,\ldots,u_{m-1},a_m,u_{m+1},\ldots,u_k} - \lambda \left( 2 \prod_{i \neq m} \varphi(u_i) \right).
\]

Then, setting the equation to be zero, we get
\[
\lambda y_u = \prod \varphi(u_i) \sum_{m=1}^{k} \prod_{i \neq m} g(u_i) \frac{1}{\varphi(a_m)} \sum_{a_m} y^{(m)}_{u_1,\ldots,u_{m-1},1,u_{m+1},\ldots,u_k}
\]

and, by a slight modification of (7.7), we get
\[
= \prod \varphi(u_i) \sum_{m=1}^{k} \prod_{i \neq m} g(u_i) \frac{1}{\varphi(u_m)} y^{(m)}_{u_1,\ldots,u_{m-1},1,u_{m+1},\ldots,u_k}.
\]

Note also that
\[
\sum_{i=1}^{k} \prod_{i \neq m} g(u_i) = \prod_{i=1}^{k} g(u_i)
\]

and substituting this in gives us a condition on \( y_u \):
\[
\lambda y_u = \prod_{i=1}^{k} \varphi(u_i) \sum_{m=1}^{k} \frac{g(u_m)}{\varphi(u_m)} y^{(m)}_{u_1,\ldots,u_{m-1},1,u_{m+1},\ldots,u_k}.
\] (7.9)
8 Smooth choice of \( y \)

As a result of the previous section we have the condition

\[
\lambda y_r = \left( \prod_{i=1}^{k} \frac{\varphi(r_i)}{g(r_i)} \right) \sum_{m=1}^{k} \frac{g(r_m)}{\varphi(r_m)} y_{r_1, \ldots, r_{m-1}, 1, r_{m+1}, \ldots, r_k}
\]

but this is very easily simplified by the condition we made in Section 5.2: that all our weights would be coprime to \( W \).

First, it is quite intuitive that \( g(x) \) will be ‘close’ to \( x \) for \( x \) with only large prime factors \( p_i \) because \( \frac{p_i-1}{p_i} \to 1 \) as \( p_i \) increases. In fact, \( \varphi(x) \) will be even closer to \( x \) in light of the fact that for any repeated prime \( p \) in the factorisation of \( x \), \( g(p^k) \) contributes \( (p-1)^k \) to \( g(x) \) whereas \( \varphi(p^k) \) provides a factor of \( p^{k-1}(p-1) \) (and for the squarefree factors of \( x \) we have \( \varphi(x) = \prod \varphi(p_i) = \prod (p_i - 1) = g(x) \)) meaning we can make the estimation \( \frac{\varphi(x)}{g(x)} \approx 1 \) and thus

\[
\lambda y_r \approx \sum_{m=1}^{k} y_{r_1, \ldots, r_{m-1}, 1, r_{m+1}, \ldots, r_k}^{(m)}.
\] \hspace{1cm} (8.1)

Before we studied the \( y_r \), it would have been reasonable to suggest that solutions might depend on the prime factorisation of \( y_r \). What the above equation gives us, is that this is not the case, and instead we have transformed our number theoretic problem into something that looks like a functional analysis question. This incentivises us to try using a smooth\(^2\) function as our \( y_r \) and see how it works out. In this case, we choose

\[
y_r = F \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right).
\] \hspace{1cm} (8.2)

Of course, for this statement to stand a chance of being useful, our function \( F \) has to take into account all the requirements we placed on our \( y_r \), from Proposition 1. Hence we set \( F \) to be 0 whenever:

1. \( r = \prod_{i=1}^{k} r_k \) satisfies either \( (r, W) > 1 \) or \( \mu(r) \neq 1 \).

2. \( \sum_{i=1}^{k} r_i > 1 \) (i.e. we need \( r \) in the unit ball of \( \mathbb{R}^k \) in the \( \| \cdot \|_1 \)-norm).

\(^2\)The word ‘smooth’ is very oversubscribed in mathematics. In number theory it describes the prime factors of a number, but here, we’ve moved on from number theory to functional analysis. Now we use ‘smooth’ to mean having continuous derivative almost everywhere.
We impose the second condition because we plan to consider when $R$ is very large and it’s neater to remove the dependence on $R$ in the domain. It is then done by scaling by $\log R$ every argument of $F$ which makes the domain of $F$ $[0, 1]^k$ (recall from Section 5.2 we have $\prod r_i < R$).

Finally, we can reach the conclusion that the last two sections have been leading us towards: the evaluations of $S_1$ and $S_2$ given in Proposition 1. We’ll begin with $S_1$.

**Lemma 8.1.** Let $y_\perp$ be given as in (8.2) for a smooth function $F$. Let

$$F_{\max} = \sup_{(t_1, \ldots, t_k) \in [0,1]^k} |F(t_1, \ldots, t_k)| + \sum_{i=1}^k \left| \frac{\partial F}{\partial t_i}(t_1, \ldots, t_k) \right|.$$ 

Then

$$S_1 = \frac{\varphi(W)^k N (\log R)^k}{W^{k+1}} I_k(F) + O \left( \frac{F_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right),$$

where, as in Proposition 1

$$I_k(F): = \int_0^1 \cdots \int_0^1 F(t_1, \ldots, t_k)^2 dt_1 \cdots dt_k.$$ 

**Proof.** From Lemma 6.1 we have

$$S_1 = \frac{N}{W} \sum_{r_1, \ldots, r_k} \frac{y_{r_1, \ldots, r_k}^2}{\prod_{i=1}^k \varphi(r_i)} + O \left( \frac{y_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right).$$

We want to substitute in our $F$, but note that the first restriction on $F$ means we need $\mu(r) \neq 1$, which is equivalent to

$$(r_i, r_j) = 1 \text{ for all } i \neq j, \quad \prod_{i=1}^k \mu(r_i)^2 = 1 \quad \text{and } (r_i, W) = 1 \text{ for all } i$$

so we add these restrictions to the sum

$$S_1 = \frac{N}{W} \sum_{r_1, \ldots, r_k \atop (r_i, W) = 1, \forall i \atop (r_i, r_j) = 1, \forall i \neq j} \prod_{i=1}^k \mu(r_i)^2 F \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)$$

$$+ O \left( \frac{F_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right).$$
To make this a little more manageable we remove the \((r_i, r_j) = 1, \forall i \neq j\) restriction and add an error term to account for that, which we will find to be of \(O(\ldots)\) smaller than our current error:

Note that if \(r_i\) and \(r_j\) share a common prime factor \(p\), then \(p > D_0\), as \((r_i, W) = (r_j, W) = 1\), giving error

\[
E := \frac{N}{W} \sum_{p > D_0} \sum_{r_1, \ldots, r_k < R \atop (r_i, W) = 1, \forall i} \prod_{i=1}^k \frac{\mu(r_i)^2}{\varphi(r_i)} F \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)^2
\]

\[
\ll \frac{N}{W} \sum_{p > D_0} \sum_{r_1, \ldots, r_k < R \atop (r_i, W) = 1, \forall i} \prod_{i=1}^k \frac{\mu(r_i)^2}{\varphi(r_i)} F^2_{\max}
\]

\[
\ll \frac{F^2_{\max} N}{W} \sum_{p > D_0} \sum_{r_1, \ldots, r_k < R \atop (r_i, W) = 1, \forall i} \prod_{i=1}^k \frac{\mu(r_i)^2}{\varphi(r_i)}
\]

Bounding this takes two steps: first we remove the dependence on \(p\) in the second sum, then we will take the product outside of the second sum.

We have the following,

\[
\sum_{u < R \atop p \mid u} \frac{1}{\varphi(u)} = \frac{1}{p} \left( \sum_{u < R} \frac{1}{\varphi(u)} + O(1) \right)
\]

\[
\leq \frac{1}{p - 1} \sum_{u < R} \frac{1}{\varphi(u)}
\]

and this translates directly over, except we will have a factor \(\frac{1}{(p-1)^2}\) because we have \(p\) dividing both \(r_i\) and \(r_j\) so each contributes \(\frac{1}{p-1}\).

By uniformity of \(r_i\), we can take the product outside the sum. As for previous sum and product formulae, we will show this for the simplest case and note that it extends easily by induction. Take

\[
\sum_{r_1, r_2} \prod_{i=1}^2 \frac{\mu(r_i)^2}{\varphi(r_i)} = \sum_{r_1} \sum_{r_2} \prod_{i=1}^2 \frac{\mu(r_i)^2}{\varphi(r_i)}.
\]
Fix $r_1$, then every term in the sum over $r_2$ has a factor of $\frac{\mu(r_1)^2}{\varphi(r_1)}$ which we can bring out the front of the sum
\[
= \sum_{r_1} \frac{\mu(r_1)^2}{\varphi(r_1)} \sum_{r_2} \frac{\mu(r_2)^2}{\varphi(r_2)}
= \left( \sum_{r} \frac{\mu(r)^2}{\varphi(r)} \right)^2.
\]
These two steps together give us
\[
\ll \frac{F_{\text{max}}^2 N}{W} \sum_{p > D_0} \frac{1}{(p-1)^2} \left( \sum_{r < R \atop (r,W)=1} \frac{\mu(r)^2}{\varphi(r)} \right)^k
\]
and we can bound this using (6.21) and the fact that
\[
\sum_{p \text{ prime}} \frac{1}{(p-1)^2} < 2,
\]
giving the whole error
\[
E \ll O \left( 2 \frac{\varphi(W)^k}{W^k} (\log R)^k \right)
\ll O \left( \frac{F_{\text{max}}^2 \varphi(W)^k N (\log R)^k}{W^k + D_0} \right).
\]
This leaves us with
\[
\sum_{r_1,\ldots,r_k \atop (r_i,W)=1, \forall i} \prod_{i=1}^k \frac{\mu(r_i)^2}{\varphi(r_i)} \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)^2
= \frac{\varphi(W)^k (\log R)^k}{W^k} I_k(F) + O \left( \frac{F_{\text{max}}^2 \varphi(W)^k (\log D_0) (\log R)^{k-1}}{W^k} \right).
\]
Proof. Take
\[ \gamma(p) = \begin{cases} 1, & p \nmid W \\ 0, & p \mid W \end{cases} \]
which is multiplicative. Then
\[ 0 \leq \frac{\gamma(p)}{p} \leq 1 - \frac{1}{2} \]
and
\[ 0 \leq \sum_{w \leq p \leq z} \frac{\gamma(p) \log p}{p} = \sum_{w \leq p \leq z} \frac{\log p}{p} \]
\[ \ll \sum_{p \mid W} \frac{\log p}{p} \]
\[ \ll \log(D_0), \]
so it satisfies the conditions of Lemma 6.2 with \( L = \log(D_0) \). Recall also from the lemma, the totally multiplicative function
\[ \eta(p) := \frac{\gamma(p)}{p - \gamma(p)} \]
and the auxiliary function
\[ P := \prod_{p \text{ prime}} \left( 1 - \frac{\gamma(p)}{p} \right)^{-1} \left( 1 - \frac{1}{p} \right). \]
Then the lemma says
\[ \sum_{d < R} \mu(d)^2 \eta(d) F \left( \frac{\log d}{\log R} \right) = P \log(R) \int_0^1 F(x)dx + O(PLF_{\text{max}}) \]
\[ = P \log(R) \int_0^1 F(x)dx + O(PLF_{\text{max}}). \]
Checking back to (6.19) and (6.20), we see that
\[ \sum_{d < R} \eta(d) = \sum_{d < R} \frac{1}{\varphi(d)} \]
and

\[ P = \frac{\varphi(W)}{W}. \]

Now we take out each \( r_i \) in turn to get (8.3) into the form we want:

\[
\sum_{r_1, \ldots, r_k} \prod_{i=1}^k \mu(r_i)^2 \varphi(r_i) \sum_{r_1} \int_0^1 \left( \varphi\left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)^2 \right) dx
\]

\[ + \sum_{\substack{r_2, \ldots, r_k \\ (r_i, r_j) = 1, \forall i \neq j}} \prod_{i=2}^k \mu(r_i)^2 \varphi(r_i) \left( \log(R) \right)^2 \]

\[ = \frac{\varphi(W)}{W} \log(R) \sum_{r_2, \ldots, r_k} \prod_{i=2}^k \mu(r_i)^2 \varphi(r_i) \left( \int_0^1 \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)^2 dx \right) \]

\[ + \sum_{\substack{r_2, \ldots, r_k \\ (r_i, r_j) = 1, \forall i \neq j}} \prod_{i=2}^k \mu(r_i)^2 \varphi(r_i) \left( \log(R) \right)^2 \]

We are left with a sum over \( k - 1 \) variables with a smooth function and we repeat \( k \) times to get

\[ \frac{\varphi(W)^k \log(R)^k}{W^k} I_k(F) \]
as our main term and we have error term
\[
\sum_{r_1, \ldots, r_{k-1}} \prod_{i=2}^{k} \frac{\mu(r_i)^2}{\varphi(r_i)} \left( O \left( \frac{\varphi(W)}{W} LF_{\max}^2 \right) \right)
\]
\[
\leq \sum_{r_1, \ldots, r_{k-1}} \prod_{i=2}^{k} \frac{\mu(r_i)^2}{\varphi(r_i)} \left( O \left( \frac{\varphi(W)}{W} LF_{\max}^2 \right) \right)
\]
\[
= \frac{\varphi(W)^{k-1}}{W^{k-1}} (\log R)^{k-1} \left( O \left( \frac{\varphi(W)}{W} LF_{\max}^2 \right) \right)
\]
\[
= O \left( \frac{F_{\text{max}}^2 \varphi(W)^k (\log D_0) (\log R)^{k-1}}{W^k} \right)
\].

Note that the error for every subsequent iteration will be smaller, so this is the only one we need to worry about.

\[\square\]

Substituting in, we get
\[
S_1 = \frac{N}{W} \varphi(W)^k \log(R)^k I_k(F) + O \left( \frac{F_{\text{max}}^2 \varphi(W)^k N (\log D_0) (\log R)^{k-1}}{W^{k+1}} \right)
\]  \hspace{1cm} (8.5)
\[
+ O \left( \frac{F_{\text{max}}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right).
\]

Then
\[
\frac{\log(R)}{D_0} \geq \log(D_0)
\]
gives the result
\[
S_1 = \frac{N}{W} \varphi(W)^k \log(R)^k I_k(F) + O \left( \frac{F_{\text{max}}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right).
\]

\[\square\]

Now we move on to the corresponding result for \(S_2\) and, as in Section 6, we will skip over the parts that are analogous to the \(S_1\) case.
Lemma 8.3. For \( y_{\mathbb{E}}, F, F_{\max} \) as in the previous lemma, we have

\[
S_2^{(m)} = \frac{\varphi(W)^{k}N(\log R)^{k+1}}{W^{k+1}\log N} J_k^{(m)}(F) + O \left( \frac{F_{\max}^2\varphi(W)^{k}N(\log R)^{k}}{W^{k+1}D_0} \right)
\]

with

\[
J_k^{(m)}(F) := \int_0^1 \cdots \int_0^1 \left( \int_0^1 F(t_1, \ldots, t_k) \, dt_m \right)^2 \, dt_1 \cdots dt_{m-1} dt_{m+1} \cdots dt_k.
\]

Proof. From Section 7, whenever we have

\[
y_{r_1, \ldots, r_k}^{(m)} = \sum_{a_m} \frac{y_{r_1, \ldots, r_{m-1}, a_m, r_{m+1}, \ldots, r_k}}{\varphi(a_m)} + O \left( \frac{y_{\max}\varphi(W)\log R}{WD_0} \right),
\]

whenever \( r_m = 1 \) and \( r = \prod_{i=1}^k r_i \) satisfies both \( (r, W) = 1 \) and \( \mu(r) \neq 0 \) and that \( y_{r_1, \ldots, r_k}^{(m)} = 0 \) otherwise. We substitute for \( F \) in terms of \( y_{\mathbb{E}} \), calling

\[
F_{x} = F \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_{m-1}}{\log R}, \frac{\log a_m}{\log R}, \frac{\log r_{m+1}}{\log R}, \ldots, \frac{\log r_k}{\log R} \right)
\]

(8.6)

to get

\[
y_{r_1, \ldots, r_k}^{(m)} = \sum_{a_m} \frac{\mu(a_m)^2}{\varphi(a_m)} F_x + O \left( \frac{y_{\max}\varphi(W)\log R}{WD_0} \right).
\]

(8.7)

Similar to before, we apply Lemma 6.2, here taking

\[
\gamma(p) = \begin{cases} 
1, & p \nmid Wr \\
0, & p \mid Wr \nend{cases},
\]


to which Lemma 6.2 applies with \( L = \log \log N \). However, this time we’re not going to apply it \( k \) times, instead we apply it just once.

Again, to simplify the equation, let

\[
F_{x}^{(m)} := \int_0^1 F \left( \frac{\log r_1}{\log R}, \ldots, \frac{\log r_{m-1}}{\log R}, \frac{\log r_{m+1}}{\log R}, \ldots, \frac{\log r_k}{\log R} \right) dt_m.
\]

(8.8)

Then we get

\[
y_{r_1, \ldots, r_k}^{(m)} = \frac{\varphi(W)\log(R)}{W} \left( \prod_{i=1}^{k} \frac{\varphi(r_i)}{r_i} \right) F_{x}^{(m)} + O \left( \frac{y_{\max}\varphi(W)\log R}{WD_0} \right).
\]

57
Then $S_2^{(m)}$ from Lemma 6.3 goes from

$$S_2^{(m)} = \frac{N}{\varphi(W) \log N} \sum_{r_1, \ldots, r_k} (y_r^{(m)})^2 \prod_{i=1}^k g(r_i) + O \left( \frac{(y_{\max}^{(m)})^2 \varphi(W)^{k-2} N (\log R)^{k-2}}{W^{k-1} D_0} \right)$$

$$+ O \left( \frac{y_{\max}^2 N}{(\log N)^4} \right)$$

to

$$S_2^{(m)} = \frac{\varphi(W) (\log R)^2 N}{W^2 \log N} \sum_{r_1, \ldots, r_k} \prod_{i=1}^k \mu(r_i)^2 \varphi(r_i)^2 g(r_i)^2 r_i^2 \left( F_r^{(m)} \right)^2$$

$$+ O \left( \frac{(F_{\max}^{(m)})^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right), \quad (8.9)$$

being careful to add all the conditions that make $y_r^{(m)} \neq 0$ and to multiply the error term by $\left( \frac{\varphi(W) \log(R)}{W} \right)^2$, which means we can drop the second error.

As before we remove the condition $(r_i, r_j) = 1$ for all $i \neq j$ by adding an extra error term which is absorbed into the existing error.

Now we wish to evaluate the main sum:

$$\sum_{r_1, \ldots, r_k} \prod_{i=1}^k \mu(r_i)^2 \varphi(r_i)^2 g(r_i)^2 r_i^2 \left( F_r^{(m)} \right)^2 \quad (8.10)$$

$$= \sum_{r_1, \ldots, r_m, r_{m+1}, \ldots, r_k} \prod_{i \neq j} \mu(r_i)^2 \varphi(r_i)^2 g(r_i)^2 r_i^2 \left( F_r^{(m)} \right)^2,$$

which we will do explicitly because we use quite a different $\gamma(p)$ from the $S_1$ case.

Take

$$\gamma(p) = \begin{cases} 
1 - \frac{p^2 - 3p + 1}{p^3 - p^2 - 2p + 1}, & p \nmid W \\
0 & p \mid W
\end{cases}$$

which satisfies the conditions of Lemma 6.3 with $L = \log(D_0)$ as this $\gamma(p)$ is slightly smaller than for the $S_1$ case. Don’t worry about the seemingly random choice of $\gamma$; it will all work out soon.
We have, whenever \( \gamma(p) \) is non-zero,

\[
\eta(p) = \left(1 - \frac{p^2 - 3p + 1}{p^3 - p^2 - 2p + 1}\right) \left(p - \frac{p^3 - 3p^2 + p}{p^3 - p^2 - 2p + 1}\right)^{-1}
\]

and the auxiliary function

\[
P = \prod_{p \text{ prime} \atop p \nmid W} \left(1 - \frac{1}{p} + \frac{p^2 - 3p + 1}{p^3 - p^2 - 2p + 1}\right)^{-1} \left(1 - \frac{1}{p}\right) \prod_{p \text{ prime} \atop p | W} \left(1 - \frac{1}{p}\right).
\]

Here’s where the strange choice of \( \gamma(p) \) comes to fruition; we chose it this way so that \( \eta(p) \) is the coefficient of \( \mu(d)^2 F(d) \), i.e.

\[
\eta(p) = \left(1 - \frac{p^2 - 3p + 1}{p^3 - p^2 - 2p + 1}\right) \left(p - \frac{p^3 - 3p^2 + p}{p^3 - p^2 - 2p + 1}\right)^{-1} = (p^3 - 2p^2 + p) \left(p^4 - 2p^3\right)^{-1} = \frac{(p-1)^2}{(p-2)p^2} = \frac{\varphi(p)^2}{g(p)p^2},
\]

which tells us that we can apply the lemma in exactly the same way we did for \( S_1 \), pulling out each \( r_i \) term \( (i \neq m) \) in turn.

We note that \( p^2 - 3p + 1 \) is positive for all \( p \geq 2 \), so our \( P \) is even smaller than for \( S_1 \) and we can use the same bound. Then with \( (k-1) \) different \( r_i \) variables to work over (not \( k \) as previously), we get

\[
\sum_{r_1, \ldots, r_{m-1}, r_{m+1}, \ldots, r_k \atop (r, W) = 1} \prod_{i \neq j} \frac{\mu(r_i)^2 \varphi(r_i)^2}{g(r_i)r_i^2} \left(F_{r_i}^{(m)}\right)^2 = \varphi(W)^{k-1} \frac{\log(R)^{k-1}}{W^{k-1}} J_k^{(m)}(F)
\]

and

\[
S_2^{(m)} = \frac{\varphi(W)^k N \log(R)^{k+1}}{W^{k+1} \log N} J_k^{(m)}(F) + O\left(\frac{(F_{\text{max}}^{(m)})^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0}\right),
\]

as required. \(\square\)
With the proof of this final lemma, we have concluded the proof of Proposition 1 and reached the end of the most technical part of this dissertation. Now we are in the enviable position of being able to prove all four theorems with relative ease.

We recall from Section 5.2 that Proposition 2 follows very swiftly from Proposition 1 and that the proof of Proposition 3 may be found in [11]. We shall conclude by proving each theorem in turn (following [11, Section 4]), with an eye on more recent bounds. Finally, we will briefly discuss the limitations of this method on improving the bounds of Theorems 1 and 2.

9 Finale

For the purposes of this paper, Proposition 1 is simply a tool for proving Proposition 2, which states, if

$$M_k := \sup_F \frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)}$$

and

$$r_k = \left\lceil \frac{\theta M_k}{2} \right\rceil$$

then, for an ordered admissible tuple $H = (h_1, \ldots, h_k)$,

$$\liminf_{n \to \infty} (p_{n+r_k-1} - p_n) \leq h_k - h_1.$$  

Proposition 3 in turn gives us:

1. $2 < M_5$
2. $4 < M_{105}$
3. $\log k - 2 \log \log k - 2 < M_k$ for sufficiently large $k$

9.1 Final Proofs

**Theorem 1.** Let $p_n$ be the $n^{th}$ prime, then

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \leq 600.$$
Proof. By Proposition 3, $M_{105} < 4$. By Bombieri - Vinogradov, we can take $\theta = \frac{1}{2} - \epsilon$ for $\epsilon > 0$.

Then we find, in Proposition 2,

$$r_k > \frac{\frac{1}{2} \times 4}{2} = 1$$

so

$$r_k = 2$$

and then we have

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \leq h_{105} - h_1.$$ 

This means we need to find an admissible tuple of length 105. The smallest possible such tuple was found by Clark-Jarvis and has length 600,$^3$ which gives our result.

Using work from [15] along with Maynard’s results, the Polymath project has concluded with this bound reduced to 246 [16].

Theorem 2 is the corresponding result for $\theta = 1 - \epsilon$:

**Theorem 2.** Assuming Elliott - Halberstam, we have:

$$\liminf_{n \to \infty} p_{n+1} - p_n \leq 12.$$ 

**Proof.** This is identical to the previous proof except we may now use $M_5 < 2$ to give $r_k = 2$.

The smallest admissible tuple of length 5 is $(0, 4, 6, 10, 12)$, proving this second theorem.

Polymath gives that this can in fact be taken as low as 6.

Finally, we have the general case, looking at $m$ being large (for $2 \leq m \leq 5$ the reader may like to look at [16, Theorem 1.4]).

**Theorem 3.** Let $p_n$ be the $n^{th}$ prime number, $m \in \mathbb{N}$, then we have

$$\liminf_{n \to \infty} (p_{n+m} - p_n) \ll m^3 e^{4m}.$$ 

$^3$For any readers who enjoy looking at long lists of prime $k$-tuples, this one and many others can be found at math.mit.edu/~primegaps/ which is run by Andrew Sutherland.
Proof. We’re looking for a similar proof to the two previous theorems, except now we have

\[ r_k > \left( \frac{1}{2} - \epsilon \right) (\log k - 2 \log \log k - 2) \]

which we want to be greater than \( m \). Taking \( \epsilon \) to be \( \frac{1}{k} \) gives us

\[ k \geq C m^2 e^{4m} \]

for some constant \( C \) which is independent of \( k \) and \( m \). Then to get this minimal we must take

\[ k = \left\lceil C m^2 e^{4m} \right\rceil \]

and find a corresponding \( k \)-tuple. Recall that for admissibility we require \( H = (h_1, \ldots, h_k) \) to have \( h_i \) all avoiding some residue class modulo each prime less than \( k \). This leads us towards one way of building an admissible \( k \)-tuple for any \( k \): we start at the first prime number after \( k \) and take the next \( k - 1 \) prime numbers after that; then none of our \( h_i \equiv 0 \) mod \( p \), for any \( p \leq k \) and our \( H \) is admissible. Using this construction we obtain

\[ H = (p_{\pi(k)+1}, \ldots, p_{\pi(k)+k}), \]

which has diameter \( p_{\pi(k)+k} - p_{\pi(k)+1} \ll k \log k \) by the prime number theorem. Then

\[ \liminf(p_{n+m} - p_n) \ll k \log k \]

and substituting in our value for \( k \), we get

\[ \liminf(p_{n+m} - p_n) \ll C m^2 e^{4m}. \]

Again, Polymath8b gives us a slightly improved bound, this time of \( C m e^{2m} \).

The last theorem says that the prime \( k \)-tuples conjecture is true for a positive proportion of admissible \( k \)-tuples. Though we fully expect it to be true for all admissible \( k \)-tuples, this partial result is a useful step in the right direction.

**Theorem 4.** Let \( A = \{a_1, a_2, \ldots, a_r\} \) be an admissible set of distinct integers and \( H = \{h_1, h_2, \ldots, h_k\} \) be a subset of \( A \) containing \( k < r \) elements. Then
for a positive proportion of such \( H, \{n + h_i\}_{i=1}^k \) are all prime for infinitely many \( n \).

Equivalently,

\[
\frac{|\{H \subseteq \mathcal{A} : H \text{ satisfies the prime } k\text{-tuples conjecture}\}|}{|\{H \subseteq \mathcal{A}\}|} \gg 1.
\]

As the statement of Theorem 4 is different from that given by Maynard, this proof is also quite different to the one he provides.

**Proof.** Given \( m \), let \( k = \lceil Cm^2 e^{4m} \rceil \) as above. First take \( r = k \) and let \( \mathcal{A} = \{a_1, a_2, \ldots, a_r\} \) be an admissible set of distinct integers. Then by Theorem 3, there exists a subset \( H = \{h_1, h_2, \ldots, h_m\} \) of \( \mathcal{A} \) such that all \( \{n + h_i\}_{i=1}^m \) are prime infinitely often. Explicitly then, we have that at least 1 of the \( 2^r \) subsets of \( \mathcal{A} \) satisfies the prime \( m \)-tuples conjecture. What we must show is that this proportion stays positive as \( r \) grows.

**Claim.** For any \( r > k, \mathcal{A} = \{a_1, a_2, \ldots, a_r\} \), the proportion of subsets of \( \mathcal{A} \) which satisfy the prime \( m \)-tuples conjecture is greater than or equal to

\[
\frac{|\{B \subseteq \mathcal{A} : |B| = k\}|}{|\{B \subseteq \mathcal{A}\}|}.
\]

**Proof.** Let \( \mathcal{A}_1 = \{a_1, a_2, \ldots, a_{r+1}\} \) be admissible. Then for each \( h_i \) in \( H \) we may take \( \mathcal{A}_1 \setminus \{h_i\} \) to obtain \( m \) admissible sets, each comprised of \( r \) integers, each possessing a different set \( H_i \) which satisfies the prime \( m \)-tuples conjecture.

The claim then follows by induction on the size of \( \mathcal{A}_1 \).

To get the theorem we have to bound this fraction. Each part can be evaluated with basic counting arguments, which give

\[
|\{B \subseteq \mathcal{A} : |B| = k\}| = \binom{r}{k},
\]

where the right hand side is the binomial coefficient, and

\[
|\{B \subseteq \mathcal{A}\}| = 2^r.
\]

This means our proportion has become

\[
\frac{r!}{2^r k!(r - k)!}.
\]
As $r$ tends to infinity the $r!$ and $(r - m)!$ terms dominate, bringing our proportion closer to 1 and giving the result

\[
\frac{|\{B \subseteq A : |B| = k\}|}{|\{B \subseteq A\}|} \gg 1.
\]

This paper has been a demonstration of the power of sieve methods, but there is a huge obstacle in the way of further progress (reducing the limit below $\lim \inf_{n \to \infty} (p_{n+1} - p_n) \leq 6$), known as the parity problem. The parity problem was first described by Selberg himself in 1949 [17], and it basically says that sieves cannot distinguish between integers with an even or odd number of prime factors. In our case, that means they cannot distinguish between primes and ‘almost-primes’, numbers that are the products of two primes. This is perhaps best emphasised by Chen’s Theorem [1], which states that the Twin Prime Conjecture, and even the more general de Polignac’s conjecture, are true if we loosen the conditions to include almost-primes. What this means for us is that if, using sieve theory, we could get any bound that said $\lim \inf_{n \to \infty} (p_{n+1} - p_n) < 6$, then we would also be able to prove certain completely contradictory statements. A more detailed version of this, including a heuristic of the contradictory statements can be seen in [16, Section 8]. It suffices to say that any improvement, in particular any proof of the Twin Prime Conjecture using sieve methods, will require a new, considerable breakthrough in sieve theory.

9.2 Acknowledgements

I would like to thank my supervisor Jennifer Balakrishnan for all her time and effort in helping me put this dissertation together and James Maynard for a very useful maths lesson.
References


